The Annals of Probability
1974, Vol. 2, No. 5, 749–791

THE 1973 WALD MEMORIAL LECTURES

THE PROBABILITY THEORY OF ADDITIVE ARITHMETIC FUNCTIONS[1,2]

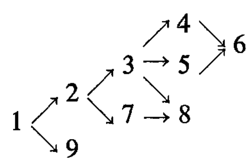
BY PATRICK BILLINGSLEY

The University of Chicago

Probability limit theorems for additive functions of positive integers are reviewed.

TABLE OF CONTENTS

A review of probability methods in multiplicative arithmetic, with liberal use of probabilistic language, reasoning, and technique, and sparing use of the apparatus of number theory, is the object of this paper. Except for the fundamental theorem of arithmetic and its immediate consequences, what number theory is required is developed in Section 10. As for probability, the ordinary limit theory for random variables is enough except for the weak-convergence results in Sections 4 and 6 and the entropy calculations in Section 9. The diagram below shows the logical interdependence of the first nine sections; the remaining three can be consulted as the need arises.



1. Introduction. *Additive functions.* An arithmetic function f, defined over the positive integers, is *additive* if $f(m_1 m_2) = f(m_1) + f(m_2)$ whenever m_1 and m_2

Received December 10, 1973; revised April 1, 1974.

[1] This paper was presented at the IMS Annual Meeting at New York, December 27–30, 1973.

[2] Supported by NSF GP 32037X and ONR N00014-67-A-0285-0009.

AMS 1970 *subject classifications.* Primary 10K20; Secondary 60F05.

Key words and phrases. Additive arithmetic function, limit theorem, distribution, entropy.

are relatively prime; it is *multiplicative* if $f(m_1 m_2) = f(m_1)f(m_2)$ whenever m_1 and m_2 are relatively prime. If, in addition, $f(p^\beta) = f(p)$ for every prime power p^β, $\beta = 1, 2, \cdots$, then f is *completely* additive or multiplicative. We shall be concerned exclusively with real arithmetic functions, and largely with additive ones, as the logarithm of a positive multiplicative function is additive.

An additive f must satisfy $f(1) = 0$. Let $\beta_p(m)$ be the exact power of the prime p in the factorization of m, so that

$$m = \Pi_p p^{\beta_p(m)} . \tag{1.1}$$

Additive functions are those of the form

$$f(m) = \sum_p f(p^{\beta_p(m)}) . \tag{1.2}$$

Let $\delta_p(m)$ be 1 or 0 according as $p \mid m$ or $p \nmid m$, so that

$$\begin{aligned} \delta_p(m) &= 0 \quad \text{if} \quad \beta_p(m) = 0 \\ &= 1 \quad \text{if} \quad \beta_p(m) > 0 . \end{aligned} \tag{1.3}$$

Completely additive functions are those of the form

$$f(m) = \sum_p f(p)\delta_p(m) . \tag{1.4}$$

The series (1.2) and (1.4) are actually finite sums, as $\beta_p(m) = \delta_p(m) = 0$ for $p > m$.

EXAMPLE 1. The number of distinct prime divisors of m is $\omega(m) = \sum_p \delta_p(m)$, which has the form (1.4) with $f(p) = 1$. With multiplicity counted, the number of prime divisors of m is $\Omega(m) = \sum_p \beta_p(m)$, which has the form (1.2) with $f(p^\beta) = \beta$. Whereas ω is completely additive, Ω is only additive.

EXAMPLE 2. If m_1 and m_2 are relatively prime, then the general divisor of $m_1 m_2$ is $k_1 k_2$, with k_1 a divisor of m_1 and k_2 a divisor of m_2. The number $\Delta(m)$ of divisors (not necessarily prime) of m is therefore multiplicative, and $\log \Delta(m)$ is additive.

EXAMPLE 3. It follows from the same observation that the sum $\sigma(m)$ of the divisors of m is multiplicative and $\log \sigma(m)$ additive.

EXAMPLE 4. Euler's function $\phi(m)$ is defined as the number of integers less than m and relatively prime to it. Because of the relation

$$\frac{\phi(m)}{m} = \Pi'_{p|m}\left(1 - \frac{1}{p}\right) \tag{1.5}$$

(see Section 7), $\phi(m)/m$ is completely multiplicative, and $\phi(m)$ is multiplicative.

EXAMPLE 5. The number of prime factors of m with multiplicity counted is at least as great as the number of distinct prime factors of m. The excess, $\Omega(m) - \omega(m)$, as the difference of two additive functions, is itself additive.

Because of the representations (1.2) and (1.4), whole classes of additive functions can be constructed at will, and these will be the objects of the general theory.

Probability. Since the value $f(m)$ of an additive function varies in a highly irregular way as m runs through the integers, it is natural to investigate the average behavior of f. Let $\mathbb{P}_n$ be the probability measure on the integers that corresponds to a mass of $1/n$ at each m in the range $1 \leqq m \leqq n$. Among the first n integers, the proportion that lie in a set M is thus $\mathbb{P}_n(M)$. The problem is to analyse, for large n, the distribution function $\mathbb{P}_n[m: f(m) \leqq x]$ of f under $\mathbb{P}_n$.

If k is a positive integer, the number of multiples of k not exceeding n is the integral part $[n/k]$ of n/k. Since $x - 1 < [x] \leqq x$, we have

$$(1.6)\qquad \frac{1}{k} - \frac{1}{n} < \mathbb{P}_n[m: k \mid m] = \frac{1}{n}\left[\frac{n}{k}\right] \leqq \frac{1}{k}.$$

For large n, the probability under $\mathbb{P}_n$ of the event $k \mid m$ is thus close to $1/k$. It is a corollary of the fundamental theorem of arithmetic that, if $p_1, \cdots, p_u$ are distinct primes, then $\delta_{p_i}(m) = 1$ for $i = 1, \cdots, u$—that is, each p_i divides m—if and only if the product $p_1 \cdots p_u$ divides m. Therefore

$$(1.7)\qquad \mathbb{P}_n[m: \delta_{p_i}(m) = 1, i = 1, \cdots, u] = \frac{1}{n}\left[\frac{n}{p_1 \cdots p_u}\right].$$

If m is drawn at random from the range $1 \leqq m \leqq n$, then the $\delta_p(m)$ are random variables, and (1.7) gives their joint distribution.

The essential point is that the $\delta_p(m)$ are almost statistically independent of one another. Consider independent random variables d_p (defined on some probability space, one variable for each prime p) satisfying

$$(1.8)\qquad \mathbb{P}[d_p = 1] = \frac{1}{p}, \qquad \mathbb{P}[d_p = 0] = 1 - \frac{1}{p}.$$

If $p_1, \cdots, p_u$ are distinct primes, then, by independence,

$$(1.9)\qquad \mathbb{P}[d_{p_i} = 1, i = 1, \cdots, u] = \frac{1}{p_1 \cdots p_u}.$$

As $n \to \infty$, (1.7) converges to (1.9) for fixed $p_1, \cdots, p_u$, so the $\delta_p(m)$ are for large n jointly distributed approximately as the simpler, independent random variables d_p are. Since, for example, $\mathbb{P}_n[m: \delta_p(m) = 1] = 0$ for $p > n$, this requires some qualification. We shall see, however, that the joint distribution under $\mathbb{P}_n$ of the $\delta_p(m)$ for $p \leqq n$ is well approximated by the joint distribution of the d_p for $p \leqq n$—sufficiently well at any rate that in many interesting cases the distribution under $\mathbb{P}_n$ of the sum (1.4) nearly coincides with the distribution of the corresponding sum $\sum_{p \leqq n} f(p) d_p$. There are of course many limit theorems concerning sums of the latter sort, and the objective is to carry them over to the analysis of completely additive arithmetic functions.

For the study of additive functions in general, we need independent, integer-valued random variables b_p having the geometric, or Pascal distribution with parameter $1/p$:

$$\mathbb{P}[b_p = k] = \left(\frac{1}{p}\right)^k \left(1 - \frac{1}{p}\right), \qquad k = 0, 1, \cdots \tag{1.10}$$

In order to have them defined on the same space, it will be convenient to define the d_p as functions of the b_p:

$$\begin{aligned} d_p &= 0 \quad \text{if} \quad b_p = 0 \\ &= 1 \quad \text{if} \quad b_p > 0 \,. \end{aligned} \tag{1.11}$$

By (1.10), this is consistent with (1.8). Compare (1.11) and (1.3).

Now $\beta_{p_i}(m) \geqq k_i$ if and only if $p_i^{k_i} \mid m$; this holds for each of distinct primes $p_1, \cdots, p_u$ if and only if $p_1^{k_1} \cdots p_u^{k_u} \mid m$. Therefore, by (1.6),

$$\mathbb{P}_n[m : \beta_{p_i}(m) \geqq k_i, i = 1, \cdots, u] = \frac{1}{n}\left[\frac{n}{p_1^{k_1} \cdots p_u^{k_u}}\right]. \tag{1.12}$$

This converges to

$$\mathbb{P}[b_{p_i} \geqq k, i = 1, \cdots, u] = \frac{1}{p_1^{k_1} \cdots p_u^{k_u}}. \tag{1.13}$$

The inclusion-exclusion principle now implies that as $n \to \infty$

$$\mathbb{P}_n[m : \beta_{p_i}(m) = k_i, i = 1, \cdots, u] \to \mathbb{P}[b_{p_i} = k_i, i = 1, \cdots, u] \tag{1.14}$$

for distinct primes p_i and nonnegative integers k_i.

The behavior of the b_p thus approximates that of the $\beta_p(m)$, and so probability theorems can be used to analyse (1.2). It will turn out that the law of large numbers and the central limit theorem apply to Examples 1 and 2 above, and the three-series theorem applies to Examples 3, 4, and 5.

Formulas concerning primes. Standard number-theoretic formulas we shall require ((1.19), (1.20), and (1.21)) can be derived heuristically by entropy considerations (which can be passed over, as there are rigorous proofs in Section 10; but see the notational conventions starting at (1.23)).

It takes $\log n$ units of information to specify an integer m in the range $1 \leqq m \leqq n$ and hence $\log n$ units of information to specify its factorization into primes—to specify, in other words, the values of

$$\beta_p(m), \qquad p \leqq n. \tag{1.15}$$

The system (1.15) thus has entropy $\log n$ under $\mathbb{P}_n$.

Let us compute the entropy of the corresponding random variables

$$b_p, \qquad p \leqq n. \tag{1.16}$$

Since b_p has mean $p^{-1}/(1-p^{-1}) = 1/(p-1)$, it has entropy

$$(1.17)\qquad -\sum_{k=0}^{\infty}\left(\frac{1}{p}\right)^k\left(1-\frac{1}{p}\right)\log\left(\frac{1}{p}\right)^k\left(1-\frac{1}{p}\right)$$
$$= \log p\cdot \mathbb{E}[b_p] - \log\left(1-\frac{1}{p}\right) = \frac{\log p}{p-1} - \log\left(1-\frac{1}{p}\right).$$

The last term, being of order $1/p$, is small compared with the next-to-last, which is itself close to $p^{-1}\log p$. Thus b_p has approximate entropy $p^{-1}\log p$, and since the b_p are independent, the system (1.16) has approximate entropy $\sum_{p\le n} p^{-1}\log p$. This sum, because of (1.14), should for large n approximate the entropy $\log n$ of the system (1.15) under $\mathbb{P}_n$:

$$(1.18)\qquad \sum_{p\le n}\frac{\log p}{p} \approx \log n .$$

The fact is that the difference of the two sides of (1.18) is bounded, and this is still true if n is replaced by a continuous variable x:

$$(1.19)\qquad \sum_{p\le x}\frac{\log p}{p} = \log x + O(1) .$$

For a direct, elementary proof of this, see Section 10.

If we regard $\pi(x)$, the number of primes not exceeding x, as a distribution function in x, we can write $\int_2^x t^{-1}\log t\,\pi(dt) \approx \log x$. If we formally take differentials on each side, we obtain $x^{-1}\log x\cdot\pi(dx) \approx x^{-1}\,dx$ and hence $\pi(dx) \approx dx/\log x$. (Treating the integral as differentiable is a far lesser transgression than differentiating the approximate equality.) Thus the density of primes in the vicinity of x should be about $1/\log x$, which integrates approximately to $x/\log x$ (differentiate again). For the entropy approximation (1.18) to hold, primes ought therefore to come along at such a rate that $\pi(x)$ is near $x/\log x$. According to the prime number theorem, the ratio of these two quantities in fact converges to 1 as $x\to\infty$. Here we shall need only the elementary one-sided estimate (see Section 10)

$$(1.20)\qquad \pi(x) = O\left(\frac{x}{\log x}\right).$$

If $\pi(dx)\approx dx/\log x$, then $\log\log x \approx \int_2^x t^{-1}\log^{-1} t\,dt \approx \int_2^x t^{-1}\pi(dt) = \sum_{p\le x} p^{-1}$. The last arithmetic fact we need (see Section 10) is that there exists a constant c such that

$$(1.21)\qquad \sum_{p\le x}\frac{1}{p} = \log\log x + c + O\left(\frac{1}{\log x}\right).$$

In Section 9 these ideas are used to give a rigorous information-theoretic proof that there are infinitely many primes. It would be interesting to have rigorous information-theoretic proofs of results like (1.19).

Let us compute the uncertainty remaining in b_p when the value of d_p is known. If $d_p = 0$, then $b_p = 0$ also, and there is no uncertainty at all. Under the condition $b_p \geqq 1$ (or $d_p = 1$), the conditional distribution of $b_p - 1$ coincides (see (1.10)) with the unconditional distribution of b_p and hence has entropy (1.17). Multiplying by $\mathbb{P}[d_p = 1] = 1/p$, we see that the conditional entropy of b_p given d_p is

$$\frac{\log p}{p(p-1)} - \frac{1}{p}\log\left(1 - \frac{1}{p}\right) = O\left(\frac{\log p}{p^2}\right). \tag{1.22}$$

Since $\sum_p p^{-2}\log p < \infty$, it follows by independence that there exists a finite K such that the entropy remaining in the system (1.16) when the d_p, $p \leqq n$, are known is less than K.

In Section 9 the number-theoretic analogue is proved: There exists a K such that, under $\mathbb{P}_n$, the conditional entropy of the system (1.15) given the system $\delta_p(m)$, $p \leqq n$, is less than K; K can be taken as 2.6—or as 3.8 if information is measured in bits rather than in natural units. Thus to know which primes divide an integer m is almost to know m itself, in that to learn the exact powers with which the various primes divide m yields less than 4 further bits of information, which for large n (m being a random integer in the range $1 \leqq m \leqq n$) is small compared with the total information of $\log_2 n$ bits. It is a reflection of this fact that for many problems an understanding of completely additive functions is enough for an understanding of additive functions in general (see Lemma 2.2 below).

Notation: *Probability*. In addition to the d_p and b_p defined above, we shall require random variables

$$d_p' = d_p - \frac{1}{p}, \qquad b_p' = b_p - \frac{1}{p}. \tag{1.23}$$

Since $\mathbb{E}[d_p] = p^{-1}$, d_p' is d_p centered at 0; since $\mathbb{E}[b_p] = p^{-1}/(1 - p^{-1})$ is nearly p^{-1}, b_p' is nearly centered at 0.

For distribution functions F_n and F, $F_n \Rightarrow F$ will denote weak convergence: $F_n(x) \to F(x)$ for continuity points x of F. For random variables ζ_n, $\zeta_n \Rightarrow F$ will denote convergence in distribution: the distribution function of ζ_n converges weakly to F; and $\zeta_n \Rightarrow \zeta$ will indicate that the distribution function of ζ_n converges weakly to that of ζ. Finally, if a is a constant, $\zeta_n \Rightarrow a$ will mean that ζ_n converges in probability to a. In all of these relations, each random variable involved may be defined on a different probability space.

Notation: *Number theory*. In addition to the arithmetic functions $\delta_p(m)$ and $\beta_p(m)$, we shall require the centered functions

$$\delta_p'(m) = \delta_p(m) - \frac{1}{p}, \qquad \beta_p'(m) = \beta_p(m) - \frac{1}{p}. \tag{1.24}$$

There an alphabetic parallel between the notation in (1.24) and (1.23). The number-theoretic results are most easily understood, and their analogy with probability results most clearly brought out, by the systematic use of parallel notation. An arithmetic function $f(m)$, as a function on the space of integers, is a random variable under $\mathbb{P}_n$. By the standard convention of probability theory, we usually omit the argument m; thus $\mathbb{P}_n[f \leqq x]$ is short for $\mathbb{P}_n[m : f(m) \leqq x]$. By

$$\mathbb{E}_n[f] = \frac{1}{n} \sum_{m=1}^{n} f(m) \tag{1.25}$$

we denote the expected value of f under $\mathbb{P}_n$.

Most of the theorems will be stated in terms not of a single additive function f, but in terms of a sequence or array $\{f_n\}$ of additive functions. We consider the distribution of f_n under $\mathbb{P}_n$ and investigate what happens as $n \to \infty$, which corresponds in probability theory to working with a triangular array of random variables. Even if we are interested in a single additive f, truncation arguments ordinarily lead to an associated array $\{f_n\}$, and so it is convenient to work with arrays from the outset. Also, limits such as the Poisson distribution can arise only from an array.

The distribution function of f_n under $\mathbb{P}_n$ is $\mathbb{P}_n[f_n \leqq x]$. If this converges weakly to a distribution function F, we write

$$f_n \Rightarrow F . \tag{1.26}$$

If ζ is a random variable,

$$f_n \Rightarrow \zeta \tag{1.27}$$

will indicate that (1.26) holds with the distribution function of ζ in the role of F. Finally, there is the analogous notion of convergence in probability: for a constant a,

$$f_n \Rightarrow a \tag{1.28}$$

will mean that $\mathbb{P}_n[|f_n - a| \geqq \varepsilon] \to 0$ for $\varepsilon > 0$. In (1.26), (1.27), and (1.28), the f_n need not be additive. It must be remembered that the distribution of f_n is governed by $\mathbb{P}_n$, and this is so even if f_n does not depend on n. If $f \Rightarrow F$, which means that $\mathbb{P}_n[f \leqq x] \to F(x)$ for continuity points x of F, we say that f *has distribution* F.

Density. If $\mathbb{P}_n(M) \to \theta$ for a set M of integers, we say M has density θ and write $D(M) = \theta$:

$$\mathbb{P}_n(M) \to D(M) . \tag{1.29}$$

A set of density 1 we regard as containing "practically all" integers, a set of density 0 as containing "practically none." An arithmetic function f has distribution F if and only if

$$D[m : f(m) \leqq x] = F(x) \tag{1.30}$$

for continuity points x of F.

2. The law of large numbers. Let $\{f_n\}$ be an array of additive functions. Suppose for the moment that f_n is completely additive, so that (see (1.4)) $f_n(m) = \sum_{p\leq n} f_n(p)\delta_p(m)$ for $m \leqq n$. The corresponding sum $\zeta_n = \sum_{p\leq n} f_n(p)d_p$ has mean

$$A_n = \sum_{p\leq n} \frac{f_n(p)}{p}\ ; \tag{2.1}$$

in cases of interest, its variance $\sum_{p\leq n} f_n^2(p)p^{-1}(1-p^{-1})$ is near

$$B_n^2 = \sum_{p\leq n} \frac{f_n^2(p)}{p}\ . \tag{2.2}$$

In analysing the distribution of f_n under $\mathbf{P}_n$ it is therefore natural to normalize f_n to $(f_n - A_n)/B_n$; it will turn out that this normalization is appropriate in the additive case as well as in the completely additive case.

In Sections 2 through 6 we shall consistently use the notation (2.1) and (2.2). We shall assume that $B_n > 0$ for sufficiently large n; if $f_n \equiv f$, this amounts only to the assumption that $f(p)$ does not vanish identically.

Since ζ_n, as defined above, has mean A_n and variance at most B_n^2, it follows by Chebyshev's inequality that $(\zeta_n - A_n)/\psi_n B_n \Rightarrow 0$ if $\psi_n \rightarrow \infty$. With the extra condition (2.3) to cover the case in which f_n is additive but not completely additive, the number-theoretic analogue holds.

THEOREM 2.1. *Let $\{f_n\}$ be an array of additive functions such that*

$$\sup_n \frac{|f_n(m)|}{B_n} < \infty\ , \qquad m = 1, 2, \cdots\ . \tag{2.3}$$

If $\psi_n \rightarrow \infty$, then

$$\frac{f_n - A_n}{\psi_n B_n} \Rightarrow 0\ . \tag{2.4}$$

We may restrict the n of the supremum in (2.3) to values large enough that $B_n > 0$. Notice that (2.3) is no restriction at all if f_n does not depend on n. Notice also that (2.3) is no restriction if each f_n is completely additive, since it then suffices to check it for prime m and since $|f_n(p)| \leqq pB_n$ by (2.2). Before proving the theorem, let us apply it to the first two of the examples in Section 1.

EXAMPLE 1. If $f_n = \omega$ for all n, then $A_n = B_n^2 = \sum_{p\leq n} 1/p$. By (1.21),

$$A_n = B_n^2 = \log\log n + O(1)\ , \tag{2.5}$$

and it follows by (2.4) that (if $\psi_n \rightarrow \infty$)

$$\frac{\omega - \log\log n}{\psi_n(\log\log n)^{\frac{1}{2}}} \Rightarrow 0\ . \tag{2.6}$$

To derive a consequence of (2.6), let us show that

$$\frac{\log\log m - \log\log n}{(\log\log n)^{\frac{1}{2}}} \Rightarrow 0\ . \tag{2.7}$$

For $m \leqq n$, the ratio here is at most 0; if it is less than $-\varepsilon$ and if $n^{\frac{1}{2}} \leqq m$, then $\log\log n^{\frac{1}{2}} < \log\log n - \varepsilon(\log\log n)^{\frac{1}{2}}$, which implies $\log\log n < \varepsilon^{-2}\log^2 2$. Therefore, for n beyond some $n_0(\varepsilon)$, the ratio in (2.7) has modulus exceeding ε with $\mathbb{P}_n$-probability at most $\mathbb{P}_n[m\colon m \leqq n^{\frac{1}{2}}]$, which goes to 0. This proves (2.7), which in turn implies $\log\log m/\log\log n \Rightarrow 1$. By this and (2.6),

$$\frac{\omega(m) - \log\log m}{(\log\log m)^{\frac{1}{2}+\varepsilon}} \Rightarrow 0$$

for $\varepsilon > 0$. Therefore (see (1.29))

$$D[m\colon |\omega(m) - \log\log m| < (\log\log m)^{\frac{1}{2}+\varepsilon}] = 1\,, \tag{2.8}$$

and $\omega(m)$ is "usually" near $\log\log m$.

These results are also true with $\Omega(m)$ in place of $\omega(m)$, since Theorem 2.1 applies as well in the additive as in the completely additive case. An integer m is "round" if $\Omega(m)$ is large—if m is the product of many rather small primes. Hardy and Wright ([19] page 358) observe that round numbers (like $1200 = 2^4 \cdot 3 \cdot 5^2$ and $2187 = 3^7$, decimal notation obscuring the roundness of the latter) are rare, as "may be verified by any one who will make a habit of factorizing numbers which, like numbers of taxi-cabs or railway carriages, are presented to his attention in a random manner." Since $\log\log 10^7 < 3$, an integer under 10 million will ordinarily have at most three prime factors. This much the law of large numbers yields; the central limit theorem of the next section contains more detailed information.

EXAMPLE 2. A prime p has $\Delta(p) = 2$ divisors, and so, if $f_n(m) = \log\Delta(m)$,

$$\begin{aligned} A_n &= \log 2 \cdot \log\log n + O(1) \\ B_n^{\,2} &= \log^2 2 \cdot \log\log n + O(1)\,. \end{aligned} \tag{2.9}$$

Therefore,

$$\frac{\log\Delta - \log 2 \cdot \log\log n}{\psi_n(\log\log n)^{\frac{1}{2}}} \Rightarrow 0 \tag{2.10}$$

if $\psi_n \to \infty$. The analogue of (2.8) follows as before, and $\log\Delta(m)$ is usually near $\log 2 \cdot \log\log m$, or $\Delta(m)$ itself is on a logarithmic scale near $(\log m)^{\log 2}$.

Let $\Delta_u(m)$ be the number of u-tuples $(k_1, \cdots, k_u)$ of positive integers with $k_1 \cdots k_u = m$—the number of ways of expressing m as a product of u factors, order accounted for. Then $\Delta_u(m)$ is multiplicative, and $\Delta(m)$ is the special case $\Delta_2(m)$. Since $\Delta_u(p) = u$, (2.10) holds with $\log\Delta_u$ and $\log u$ in place of $\log\Delta$ and $\log 2$.

The following lemma, which we shall use repeatedly, suffices to prove Theorem 2.1 for the case in which f_n is completely additive. Let g be a function defined over the primes.

LEMMA 2.1. *There is a constant C, independent of g and of n, such that*

$$\mathbb{E}_n[\textstyle\sum_{p\leqq n} g(p)\delta_p']^2 \leqq C \textstyle\sum_{p\leqq n} \dfrac{g^2(p)}{p}\,. \tag{2.11}$$

By the Chebyshev argument, (2.11) implies

$$\mathbb{P}_n[|\sum_{p\leqq n} g(p)\delta_p'| \geqq \lambda] \leqq \frac{C}{\lambda^2} \sum_{p\leqq n} \frac{g^2(p)}{p}. \tag{2.12}$$

For completely additive f_n, $f_n(m) - A_n = \sum_{p\leqq n} f_n(p)\delta_p'(m)$ if $m \leqq n$, and hence (2.12) implies (2.4) in this case. Obviously, the probability analogues of (2.11) and (2.12) hold with $C = 1$.

PROOF OF THE LEMMA. Let $A = \sum_{p\leqq n} g(p)/p$ and $B^2 = \sum_{p\leqq n} g^2(p)/p$. We first find a C such that

$$\mathbb{E}_n[\sum_{p\leqq n} g(p)\delta_p]^2 \leqq A^2 + CB^2 \tag{2.13}$$

for all g and n. By (1.7) (here p and q both represent primes),

$$\begin{aligned} &\mathbb{E}_n[\sum_{p\leqq n} g(p)\delta_p]^2 \\ &\quad = \sum_{p\leqq n} g^2(p)\frac{1}{n}\left[\frac{n}{p}\right] + \sum_{pq\leqq n, p\neq q} g(p)g(q)\frac{1}{n}\left[\frac{n}{pq}\right] \\ &\quad \leqq B^2 + \sum_{pq\leqq n, p\neq q} \frac{g(p)g(q)}{pq} + \frac{1}{n}\sum_{pq\leqq n} |g(p)g(q)|. \end{aligned} \tag{2.14}$$

By Schwarz's inequality,

$$\sum_{pq\leqq n} |g(p)g(q)| \leqq \left[\sum_{pq\leqq n} pq \cdot \sum_{pq\leqq n} \frac{g^2(p)g^2(q)}{pq}\right]^{\frac{1}{2}} \leqq B^2[\sum_{pq\leqq n} pq]^{\frac{1}{2}}.$$

If we show that $\sum_{pq\leqq n} pq = O(n^2)$, it will follow that the final term in (2.14) is at most a bounded multiple of B^2. But $\sum_{pq\leqq n} pq \leqq n\sum_{p\leqq n} \pi(n/p)$. For $p \leqq n^{\frac{1}{2}}$, use (1.20) to bound $\pi(n/p)$ by a multiple of $np^{-1}/\log np^{-1} \leqq 2np^{-1}/\log n$; elsewhere, bound $\pi(n/p)$ by n/p. It follows that

$$\sum_{pq\leqq n} pq = O\left(\frac{n^2}{\log n}\sum_{p\leqq n^{\frac{1}{2}}} \frac{1}{p} + n^2 \sum_{n^{\frac{1}{2}}<p\leqq n} \frac{1}{p}\right).$$

and this is $O(n^2)$ by (1.21).

Since the final term in (2.14) is at most a bounded multiple of B^2, (2.13) will follow if we show that the next-to-last sum in (2.14) is at most $A^2 + C_0B^2$ for a universal C_0. This sum does not decrease if terms with $p = q$ are included, and A^2 is this latter sum with the range further expanded to include pairs $p, q \leqq n$ with $pq > n$. Thus

$$\sum_{pq\leqq n, p\neq q} \frac{g(p)g(q)}{pq} \leqq A^2 + \sum_{p,q\leqq n, pq>n} \frac{|g(p)g(q)|}{pq},$$

and for the proof of (2.13) it remains only to estimate the last sum, which by Schwarz's inequality is at most

$$\left[\sum_{p,q\leqq n, qp>n} \frac{1}{pq} \cdot \sum_{p,q\leqq n} \frac{g^2(p)g^2(q)}{pq}\right]^{\frac{1}{2}} = B^2\left[\sum_{p,q\leqq n, pq>n} \frac{1}{pq}\right]^{\frac{1}{2}}.$$

Now

$$\sum_{p,q\leqq n,pq>n} \frac{1}{pq} \leqq 2 \sum_{p\leqq n^{\frac{1}{2}},n/p<q\leqq n} \frac{1}{pq} + \left[\sum_{n^{\frac{1}{2}}<p\leqq n} \frac{1}{p}\right]^2 .$$

By (1.21), the second sum on the right is $O(1)$ and the first one is

$$\begin{aligned}\sum_{p\leqq n^{\frac{1}{2}}} \frac{1}{p}\left[\log\log n - \log\log\frac{n}{p} + O\left(\frac{1}{\log n/p}\right)\right] \\ = \sum_{p\leqq n^{\frac{1}{2}}} \frac{1}{p}\left[-\log\left(1-\frac{\log p}{\log n}\right) + O\left(\frac{1}{\log n}\right)\right] \\ = \sum_{p\leqq n^{\frac{1}{2}}} \frac{1}{p} O\left(\frac{\log p}{\log n}\right).\end{aligned}$$

The final term here is $O(1)$ by (1.19), and so (2.13) is proved.

If $A' = \mathbb{E}_n[\sum_{p\leqq n} g(p)\delta_p]$, then by Schwarz's inequality,

$$|A - A'| \leqq \frac{1}{n} \sum_{p\leqq n} |g(p)| \leqq \frac{1}{n} [\sum_{p\leqq n} p]^{\frac{1}{2}} \cdot B \leqq B\left[\frac{\pi(n)}{n}\right]^{\frac{1}{2}} .$$

By Schwarz's inequality again and (1.21),

$$|A| \leqq \sum_{p\leqq n} \frac{|g(p)|}{p} \leqq \left[\sum_{p\leqq n} \frac{1}{p}\right]^{\frac{1}{2}} \cdot B = B \cdot O(\log\log n)^{\frac{1}{2}} ,$$

and it follows by (1.20) that $|A| \cdot |A - A'|$ is at most a bounded multiple of B^2. By (2.13), $\mathbb{E}_n[\sum_{p\leqq n} g(p)\delta_p - A]^2 \leqq CB^2 + 2A(A - A')$, which implies (2.11) with some new constant C. □

To each additive f there is associated a unique completely additive f^* defined by

$$f^*(m) = \sum_p f(p)\delta_p(m) . \tag{2.15}$$

For example, $\Omega^* = \omega$. Notice that $f^* = f$ if f itself is completely additive and that $f^*(p) = f(p)$ in any case. The quantities A_n and B_n^2, as defined by (2.1) and (2.2), are the same for f^* as for f. If $\psi_n \to \infty$, then $(f_n^* - A_n)/\psi_n B_n \Rightarrow 0$ by (2.12); hence Theorem 2.1 for the general additive array will follow if we prove that $(f_n - f_n^*)/\psi_n B_n \Rightarrow 0$. This is a consequence of (2.3) and the following lemma (with $C_n = \psi_n B_n$). The lemma will be used repeatedly.

LEMMA 2.2. *If $\{f_n\}$ is an array of additive functions and C_n are positive constants with*

$$\lim_{n\to\infty} \frac{f_n(m)}{C_n} = 0 , \qquad m = 1, 2, \cdots , \tag{2.16}$$

then

$$\frac{|f_n - f_n^*|}{C_n} \leqq \frac{1}{C_n} \sum_p |f_n(p^{\beta_p}) - f_n^*(p^{\beta_p})| \Rightarrow 0 . \tag{2.17}$$

PROOF. The sum in (2.17) is at most

$$\sum_{p\leqq T} [|f_n(p^{\beta_p})| + |f(p)|] + \sum_{p>T} |f_n(p^{\beta_p}) - f_n^*(p^{\beta_p})| ,$$

and the second sum here vanishes unless $\beta_p \geqq 2$ for some $p > T$. Therefore

$$(2.18) \quad \mathbb{P}_n\left[\frac{1}{C_n} \sum_p |f_n(p^{\beta_p}) - f_n^*(p^{\beta_p})| \geqq \varepsilon\right] \leqq \sum_{p>T} \frac{1}{p^2} + \sum_{p \leqq T} \mathbb{P}_n\left[|f_n(p^{\beta_p})| + |f_n(p)| \geqq \frac{\varepsilon C_n}{T}\right].$$

Choose T so large that $\sum_{p>T} 1/p^2 < \varepsilon/2$; that T fixed, choose k so large that

$$\sum_{p \leqq T} \mathbb{P}_n[\beta_p \geqq k] \leqq \sum_{p \leqq T} \frac{1}{p^k} < \tfrac{1}{2}\varepsilon .$$

Finally, use (2.16) to choose n_0 so that $n \geqq n_0$ implies $2|f_n(m)| < \varepsilon C_n/T$ for all $m \leqq T^k$. If $n \geqq n_0$, then the right side of (2.18) is at most ε. □

The point of the lemma is this. If $\beta_p(m) \leqq 1$, then $f_n(p^{\beta_p(m)}) = f_n^*(p^{\beta_p(m)})$; hence if $\beta_p(m) \leqq 1$ for all but a few p, $f_n(m)$ should be near $f_n^*(m)$. Since $\mathbb{P}_n[\beta_p \geqq 2] \leqq 1/p^2$ and $\sum 1/p^2 < \infty$, it is likely that $\beta_p(m) \leqq 1$ for most p, and so it is likely that $f_n(m)$ is near $f_n^*(m)$. For this reason, most problems concerning additive functions reduce to problems concerning completely additive functions. See the remarks following (1.22) and at the end of Section 9.

3. Approximation by the normal law. Here we prove the central limit theorem under conditions which make the arguments quite simple. Section 5 contains the general theory of convergence to infinitely divisible laws, including the central limit theorem under the Lindeberg conditions.

Let Φ be the standard normal distribution function.

THEOREM 3.1. *If $\{f_n\}$ is an array of additive functions such that*

$$(3.1) \quad \lim_{n\to\infty} \frac{f_n(m)}{B_n} = 0 , \qquad m = 1, 2, \cdots ,$$

and

$$(3.2) \quad \max_{p \leqq n} \frac{|f_n(p)|}{B_n} \to 0 ,$$

then

$$(3.3) \quad \frac{f_n - A_n}{B_n} \Rightarrow \Phi .$$

If $f_n \equiv f$ and $B_n \to \infty$, then (3.1) automatically holds, and if, further, $\sup_p |f(p)| < \infty$, then (3.2) holds. These remarks apply to Examples 1 and 2. (For completely additive f_n, (3.2) implies (3.1).)

EXAMPLE 1. If $f_n = \omega$, then (2.5) holds, so (3.3) implies

$$\frac{\omega - \log\log n}{(\log\log n)^{\frac{1}{2}}} \Rightarrow \Phi .$$

Of course this is stronger than (2.6). Because of (2.7), we also have

$$\frac{\omega(m) - \log\log m}{(\log\log m)^{\frac{1}{2}}} \Rightarrow \Phi ,$$

which can be put in the form (see (1.29))

$$D[m: \omega(m) \leqq \log\log m + x(\log\log m)^{\frac{1}{2}}] = \Phi(x) .$$

These results hold with Ω in place of ω.

EXAMPLE 2. By the same reasoning (see (2.9)),

$$\frac{\log\Delta - \log 2 \cdot \log\log n}{\log 2 \cdot (\log\log n)^{\frac{1}{2}}} \Rightarrow \Phi ,$$

and

$$D[m: \Delta(m) \leqq (\log m)^{\log 2} 2^{x(\log\log m)^{\frac{1}{2}}}] = \Phi(x) .$$

There are obvious analogues for $\Delta_u(m)$ (defined after Example 2 in the preceding section).

To prove Theorem 3.1, observe first that, by (3.1) and Lemma 2.2, $(f_n - f_n^*)/B_n \Rightarrow 0$. We may therefore assume in the proof that f_n is completely additive, in which case (3.3) is the same thing as

$$\frac{1}{B_n} \sum_{p \leqq n} f_n(p)\delta_p' \Rightarrow \Phi . \tag{3.4}$$

The next step is to find a sequence $\{r_n\}$ in which r_n is small enough that

$$\frac{\log r_n}{\log n} \to 0 \tag{3.5}$$

but large enough that

$$\frac{1}{B_n^2} \sum_{r_n < p \leqq n} \frac{f_n^2(p)}{p} \to 0 . \tag{3.6}$$

If ε_n is the maximum in (3.2), then $\varepsilon_n \to 0$, so that $r_n = n^{\varepsilon_n}$ satisfies (3.5). The quantity in (3.6) is at most (use (1.21))

$$\begin{aligned} \varepsilon_n^2 \sum_{r_n < p \leqq n} \frac{1}{p} &= \varepsilon_n^2(\log\log n - \log\log r_n + o(1)) \\ &= \varepsilon_n^2(-\log \varepsilon_n + o(1)) , \end{aligned}$$

from which (3.6) follows.

By (3.6) and Lemma 2.1, (3.4) is equivalent to

$$\frac{1}{B_n} \sum_{p \leqq r_n} f_n(p)\delta_p' \Rightarrow \Phi . \tag{3.7}$$

Consider the probability analogue

$$\frac{1}{B_n} \sum_{p \leqq r_n} f_n(p)d_p' \Rightarrow \Phi . \tag{3.8}$$

Since

$$\sum_{p \leqq n} \frac{f_n^2(p)}{p^2} \leqq u \max_{p \leqq u} f_n^2(p) + \frac{1}{u} B_n^2 ,$$

it follows by (3.1) that the sum on the left is $o(B_n^2)$; hence $\sum_{p\leq n} f_n(p)d_p'$ has asymptotic variance B_n^2, and by (3.6), so has the sum in (3.8). Now $|d_p'| \leqq 1$, and so $\mathbb{E}[|d_p'|^3] \leqq \mathbb{E}[d_p']^2 \leqq 1/p$; hence the third absolute moment of $f_n(p)d_p'$ is at most $|f_n(p)|^3/p$. Since $B_n^{-3} \sum_{p\leq r_n} |f_n(p)|^3/p$ is at most the maximum in (3.2), Lyapounov's theorem ([7] page 185) implies (3.8).

We deduce (3.7) from (3.8) by a moment argument requiring two lemmas we shall use again further on. Let g be a function defined over a set U of primes and put

$$B^2 = \sum_{p\in U} \frac{g^2(p)}{p}\,. \tag{3.9}$$

LEMMA 3.1. *If* max $U \leqq r \leqq n$, *then for all* k

$$|\mathbb{E}_n[\sum_{p\in U} g(p)\delta_p']^k - \mathbb{E}[\sum_{p\in U} g(p)d_p']^k| \leqq \frac{2^k r^k B^k}{n}\,. \tag{3.10}$$

The inequality (3.10) is an expression of the fact that the joint distributions of the δ_p and of the d_p nearly coincide. The next result concerns the d_p alone.

LEMMA 3.2. *If* $|g(p)| \leqq M$ *for* $p \in U$, *where* $M \geqq 1$, *then for all* k

$$|\mathbb{E}[\sum_{p\in U} g(p)d_p']^k| \leqq k!\, M^k e^{kB^2}\,. \tag{3.11}$$

Deferring the proofs of the lemmas, let us use them to complete the proof of the theorem. Let h_n and ζ_n denote the nth elements of the sequences in (3.7) and (3.8). If we take $g(p) = f_n(p)/B_n$ and $U = [p : p \leqq r_n]$ in the lemmas, then (3.9) is at most 1, and (3.10) gives

$$|\mathbb{E}_n[h_n^k] - \mathbb{E}[\zeta_n^k]| \leqq \frac{2^k r_n^k}{n}\,, \tag{3.12}$$

which goes to 0 for each k because of (3.5). By (3.2), there exists an $M \geqq 1$ such that $|f_n(p)|/B_n \leqq M$ for all n and all $p \leqq n$, and so (3.11) implies

$$|\mathbb{E}[\zeta_n^k]| \leqq k!\, M^k e^k\,.$$

By Theorem 11.2 with $\theta = Me$, (3.8) implies (3.7).

The ideas in this proof will recur in Section 5. If in (3.7) and (3.8) the sums ranged over $p \leqq n$, the right side of (3.12) would be $2^k n^k/n$, which goes to infinity, not 0. Hence the necessity of restricting p to the range $p \leqq r_n$, where r_n satisfies (3.5). On the other hand, r_n must satisfy (3.6) to ensure that (3.4) and (3.7) are equivalent.

PROOF OF LEMMA 3.1. If $h = \sum_{p\in U} g(p)\delta_p$ and $\zeta = \sum_{p\in U} g(p)d_p$, then

$$\mathbb{E}_n[h^k] = \sum g(p_1)\cdots g(p_k)\mathbb{E}_n[\delta_{p_1}\cdots\delta_{p_k}] \tag{3.13}$$

and

$$\mathbb{E}[\zeta^k] = \sum g(p_1)\cdots g(p_k)\mathbb{E}[d_{p_1}\cdots d_{p_k}]\,, \tag{3.14}$$

where in each sum the p_j independently range over U. There may of course be repeats among $p_1, \cdots, p_k$; if $p_{i_1}, p_{i_2}, \cdots$ is a list of the distinct primes among $p_1, \cdots, p_k$, and if $P = p_{i_1}p_{i_2}\cdots$ is their product, then $\mathbb{E}_n[\delta_{p_1}\cdots\delta_{p_k}] = \mathbb{P}_n[\delta_{p_{i_1}} = \delta_{p_{i_2}} = \cdots = 1] = n^{-1}[n/P]$ and $\mathbb{E}[d_{p_1}\cdots d_{p_k}] = \mathbb{P}[d_{p_{i_1}} = d_{p_{i_2}} = \cdots = 1] = 1/P$. Since these two expected values differ by at most $1/n$, a term-by-term comparison of the sums (3.13) and (3.14) shows that

$$(3.15) \qquad |\mathbb{E}_n[h^k] - \mathbb{E}[\zeta^k]| \leqq \frac{1}{n}[\sum_{p\in U}|g(p)|]^k \leqq \frac{r^kB^k}{n},$$

where the last inequality holds because, by Schwarz's inequality

$$\sum_{p\in U}|g(p)| \leqq [\sum_{p\in U}p]^{\frac{1}{2}}B \leqq rB .$$

Let $A = \sum_{p\in U} g(p)/p$. Now

$$\mathbb{E}_n[h-A]^k = \sum_{i=0}^k \binom{k}{i}\mathbb{E}_n[h^i](-A)^{k-i},$$

and $\mathbb{E}[\zeta - A]^k$ has the analogous expansion. Comparing the expansions and applying (3.15) to each term gives

$$(3.16) \qquad |\mathbb{E}_n[h-A]^k - \mathbb{E}[\zeta - A]^k| \leqq \sum_{i=0}^k \binom{k}{i}\frac{r^iB^i}{n}|A|^{k-i} = \frac{1}{n}(|A| + rB)^k .$$

Now (3.10) follows because $|A| \leqq \sum_{p\in U}|g(p)| \leqq rB$. □

PROOF OF LEMMA 3.2. By the multinomial expansion,

$$(\sum_{p\in U} x_p)^k = \sum_{u=1}^k \sum{}' \frac{k!}{k_1!\cdots k_u!}\sum{}'' x_{p_1}^{k_1}\cdots x_{p_u}^{k_u},$$

where $\sum'$ extends over the u-tuples $(k_1, \cdots, k_u)$ of positive integers adding to k and $\sum''$ extends over the u-tuples $(p_1, \cdots, p_u)$ of primes in U with $p_1 < \cdots < p_u$. Therefore the kth moment in question (see (3.11)) is

$$(3.17) \qquad \sum_{u=1}^k \sum{}' \frac{k!}{k_1!\cdots k_u!}\sum{}'' g^{k_1}(p_1)\mathbb{E}[d'_{p_1}]^{k_1}\cdots g^{k_u}(p_u)\mathbb{E}[d'_{p_u}]^{k_u}.$$

Since the d_p' have mean 0, the summand in (3.17) vanishes if $k_i = 1$ for some i; and $k_i \geqq 2$ implies $|\mathbb{E}[d'_{p_i}]^{k_i}| \leqq \mathbb{E}[d'_{p_i}]^2 \leqq 1/p_i$. Therefore

$$(3.18) \qquad |\mathbb{E}[\sum_{p\in U} g(p)d_p']^k| \leqq k!\sum_{u=1}^k \sum{}'\sum{}'' \frac{|g^{k_1}(p_1)|}{p_1}\cdots\frac{|g^{k_u}(p_u)|}{p_u},$$

where now $\sum'$ extends over the $(k_1, \cdots, k_u)$ which add to k and in which each k_i is at least 2, and where $\sum''$ has the same range as before. Since

$$\sum_{p\in U}\frac{|g^{k_i}(p)|}{p} \leqq M^{k_i-2}\sum_{p\in U}\frac{g^2(p)}{p} \leqq M^{k_i}B^2,$$

the innermost sum in (3.18) cannot exceed

$$\frac{1}{u!}\prod_{i=1}^u\left[\sum_{p\in U}\frac{|g^{k_i}(p)|}{p}\right] \leqq \frac{M^kB^{2u}}{u!}.$$

Since there are fewer than k^u terms in $\sum'$, we arrive at

$$|\mathbb{E}[\sum_{p\in U} g(p)d_p']^k| \leqq k! \sum_{u=1}^{k} \sum' \frac{M^k B^{2u}}{u!} \leqq k!\, M^k \sum_{u=0}^{\infty} \frac{1}{u!} k^u B^{2u} . \qquad \square$$

4. Approximation by Brownian motion. We turn now to an invariance principle, or functional central limit theorem. The hypotheses are those of Theorem 3.1; Section 6 contains more general results. The weak-convergence theory needed here, in particular the theory of the function space $D = D[0, 1]$, can be found in [3].

For $1 \leqq m \leqq n$, define an element $X_n(\cdot, m)$ of the space D as follows. For $0 \leqq t \leqq 1$, put

$$Q_n(t) = \left[p : p \leqq n, \frac{1}{B_n^2} \sum_{q \leqq p} \frac{f_n^2(q)}{q} \leqq t\right]; \tag{4.1}$$

then define

$$X_n(t, m) = \frac{1}{B_n} \sum_{p \in Q_n(t)} \left[f_n(p^{\beta_p(m)}) - \frac{f_n(p)}{p}\right]. \tag{4.2}$$

If f_n is completely additive, (4.2) reduces to

$$X_n(t, m) = \frac{1}{B_n} \sum_{p \in Q_n(t)} f_n(p)\delta_p'(m) . \tag{4.3}$$

As t ranges over $[0, 1]$, the sums in (4.3) range over the partial sums $\sum_{p \leqq x} f_n(p)\delta_p'$, and similarly for the sums in (4.2). The scaling is so arranged that the variance sum (2.2) corresponding to $X_n(t, \cdot)$ is approximately t.

Now $X_n(\cdot, m)$ is a random function in D if m is chosen at random in the range $1 \leqq m \leqq n$. Thus we have a random element X_n of D, governed by $\mathbb{P}_n$. Let W denote Brownian motion.

THEOREM 4.1. *If $\{f_n\}$ is an array of additive functions such that*

$$\lim_{n\to\infty} \frac{f_n(m)}{B_n} = 0 , \qquad m = 1, 2, \cdots , \tag{4.4}$$

and

$$\max_{p \leqq n} \frac{|f_n(p)|}{B_n} \to 0 , \tag{4.5}$$

then

$$X_n \Rightarrow W . \tag{4.6}$$

Before proving the theorem, we apply it to ω and Δ.

EXAMPLE 1. If $f_n = \omega$, then $X_n(t, m) = B_n^{-1} \sum_{p \in Q_n(t)} \delta_p'(m)$. Let $Y_n(\cdot, m)$ be the function on $[0, 1]$ with value

$$Y_n\left(\frac{\log \log p}{\log \log n}, m\right) = \frac{1}{(\log \log n)^{\frac{1}{2}}} [\sum_{q \leqq p} \delta_q(m) - \log \log p]$$

throughout the interval from $\log\log p/\log\log n$ to the next such point to the right (and with value 0 to the left of $\log\log 2/\log\log n$). The sum $\sum_{p\leq q}\delta_q(m)$ is the number of prime factors of m not exceeding p. From $X_n \Rightarrow W$ we can deduce $Y_n \Rightarrow W$: Let $\lambda_n(0) = 0$, $\lambda_n(1) = 1$, and $\lambda_n(\log\log p/\log\log n) = B_p^2/B_n^2$, $p \leqq n$; define λ_n on $[0, 1]$ by interpolating linearly between these points. Then $\sup_t |\lambda_n(t) - t| \to 0$, and hence (see Section 17 of [3]) the random function $X_n(\lambda_n(t), m)$ converges in distribution to W. With B_n^2 and B_p^2 in place of $\log\log n$ and $\log\log p$ on the right in the definition of Y_n, it coincides with this last random function. To pass to $Y_n \Rightarrow W$ is now simple because we have $B_n^2 \sim \log\log n$ and even $\max_{p\leq n} |B_p^2 - \log\log p|/B_n \to 0$.

Since the maximum jump in Y_n goes to 0, $Y_n \Rightarrow W$ still holds if Y_n is defined as before at points of the form $\log\log p/\log\log n$ but is defined by linear interpolation in between such points. It is in this last form that Y_n is treated in [6].

Let $Z(\cdot, m)$ be the function on $[0, 1]$ with value

$$Z\left(\frac{\log\log p}{\log\log m}, m\right) = \frac{1}{(\log\log m)^{\frac{1}{2}}}\left[\textstyle\sum_{q\leq p}\delta_q(m) - \log\log p\right]$$

in the interval from $\log\log p/\log\log m$ to the next such point. Since $Y_n \Rightarrow W$ and

$$\max_{p\leq n}\left|\frac{\log\log p}{\log\log n} - \frac{\log\log p}{\log\log m}\right| \Rightarrow 0$$

(see (2.7)), it follows by the arguments of Section 17 of [3] that the P_n-distribution of Z converges to Wiener measure W. Therefore (see (1.29)) $D[m : Z(\cdot, m) \in M] = W(M)$ if M is a Borel set with $W(\partial M) = 0$. For example (see Section 11 of [3]),

$$D\left[m : \max_{p\leq m} \frac{\sum_{q\leq p}\delta_q(m) - \log\log p}{(\log\log m)^{\frac{1}{2}}} \geqq x\right] = \frac{2}{(2\pi)^{\frac{1}{2}}}\int_x^\infty e^{-u^2/2}\,du$$

for $x \geqq 0$, and

$$D\left[m : \frac{1}{\log\log m}\sum\left[\frac{1}{p} : p \leqq m,\ \textstyle\sum_{q\leq p}\delta_q(m) > \log\log p\right] \leqq x\right] = \frac{2}{\pi}\arcsin x^{\frac{1}{2}}$$

for $0 \leqq x \leqq 1$.

The same arguments go through for $f_n = \Omega$; it is only necessary to replace δ_p by β_p throughout.

EXAMPLE 2. If $f_n = \log\Delta$, then, since $\Delta(p^\beta) = \beta + 1$, (4.2) becomes

$$X_n(t, m) = \frac{1}{B_n}\sum_{p\in Q_n(t)}\left[\log(\beta_p(m) + 1) - \frac{\log 2}{p}\right].$$

The set $Q_n(t)$ is here the same as in Example 1, and by the argument there we can prove $Y_n \Rightarrow W$, where

$$Y_n\left(\frac{\log\log p}{\log\log n}, m\right) = \frac{1}{\log 2(\log\log n)^{\frac{1}{2}}} \times \left[\textstyle\sum_{q\leq p}\log(\beta_q(m) + 1) - \log 2\cdot\log\log p\right].$$

The sum on the right is the logarithm of the number of divisors of m composed entirely of primes not exceeding p. It is easy to write down the analogue of Z in the preceding example.

Because of the relation ([15] **1** page 38)

$$\Delta_u(p^\beta) = \binom{u+\beta-1}{u-1},$$

it is possible to deduce and interpret the corresponding results for Δ_u (see Example 2 in Section 2).

In order to prove Theorem 4.1, observe first that replacing f_n by $f_n{}^*$, as defined by (2.15), has no effect on (4.1) but in general converts (4.2) into a new function $X_n{}^*(t, m)$. But

$$\sup_t |X_n(t, m) - X_n{}^*(t, m)| \leqq \frac{1}{B_n} \sum_{p \leqq n} |f_n(p^{\beta_p(m)}) - f_n{}^*(p^{\beta_p(m)})| \Rightarrow 0$$

by (4.4) and Lemma 2.2. In proving (4.6) we may therefore assume that f_n is completely additive, and we may work with (4.3) rather than (4.2).

We next show that there is no loss of generality in assuming

$$f_n(p) = 0 \qquad \text{for} \quad p > n^{\frac{1}{4}} \tag{4.7}$$

(this restriction is needed in the tightness argument). Indeed, since there exists a sequence $\{r_n\}$ satisfying (3.5) and (3.6), certainly

$$\frac{1}{B_n} \sum_{n^{\frac{1}{4}} < p \leqq n} \frac{f_n^2(p)}{p} \to 0\,. \tag{4.8}$$

Hence if we pass to a new completely additive array by setting $f_n(p) = 0$ for $p > n^{\frac{1}{4}}$, the new $B_n{}^2$ is asymptotic to the old one. Since $\mathbb{E}_n[|\delta_p'|] \leqq 2/p$, it follows by the Chebyshev argument and Schwarz's inequality that

$$\mathbb{P}_n\left[\frac{1}{B_n} \sum |f_n(p)\delta_p'| \geqq \varepsilon\right] \leqq \frac{2}{\varepsilon B_n} \sum \frac{|f_n(p)|}{p}$$
$$\leqq \frac{2}{\varepsilon B_n}\left[\sum \frac{1}{p} \cdot \sum \frac{f_n^2(p)}{p}\right]^{\frac{1}{2}},$$

where each sum extends over $n^{\frac{1}{4}} < p \leqq n$. This goes to 0 by (1.21) and (4.8), so passing to the new array has in the limit no effect on the distribution of X_n.

We shall prove that $(X_n(s), X_n(t)) \Rightarrow (W_s, W_t)$ for $0 \leqq s \leqq t \leqq 1$. An obvious extension of the argument will show that all the finite-dimensional distributions converge weakly. It is enough (see [3] page 49, for example) to show that

$$aX_n(s) + bX_n(t) \Rightarrow aW_s + bW_t \tag{4.9}$$

for real a and b. If

$$\begin{aligned} g_n(p) &= (a+b)f_n(p) && \text{if} \quad p \in Q_n(s) \\ &= bf_n(p) && \text{if} \quad p \in Q_n(t) - Q_n(s) \\ &= 0 && \text{if} \quad p \notin Q_n(t)\,, \end{aligned} \tag{4.10}$$

then (4.9) is the same thing as

$$\frac{1}{B_n}\sum_{p\leq n} g_n(p)\delta_p' \Rightarrow aW_s + bW_t\,. \tag{4.11}$$

The variance sum (2.2) corresponding to g_n is

$$V_n^2 = (a+b)^2\sum_{p\in Q_n(s)}\frac{f_n^2(p)}{p} + b^2\sum_{p\in Q_n(t)-Q_n(s)}\frac{f_n^2(p)}{p}\,. \tag{4.12}$$

By the definition (4.1),

$$\sup_t\left|t - \frac{1}{B_n^2}\sum_{p\in Q_n(t)}\frac{f_n^2(p)}{p}\right| \leq \max_{p\leq n}\frac{f_n^2(p)}{B_n^2 p}\,, \tag{4.13}$$

and by (4.5),

$$\max_{p\leq n}\frac{f_n^2(p)}{B_n^2 p}\to 0\,. \tag{4.14}$$

Now (4.13) and (4.14) together imply

$$\frac{V_n^2}{B_n^2}\to (a+b)^2 s + b^2(t-s)\,. \tag{4.15}$$

Since this limit is the variance of the normal variable $aW_s + bW_t$, and since $V_n^{-1}\sum_{p\leq n} g_n(p)\delta_p' \Rightarrow \Phi$ by Theorem 3.1, (4.11) follows.

The tightness argument remains. Here the proof is constructed so as to apply in the more general setting of Section 6: *We assume only* (4.7) *and* (4.14). The argument requires two lemmas. Let $g(p)$ be a function of primes.

LEMMA 4.1. *If U_1 and U_2 are disjoint sets of primes with* $\max U_1 \leq n^{\frac{1}{4}}$ *and* $\max U_2 \leq n^{\frac{1}{4}}$, *then*

$$\begin{aligned}\mathbb{E}_n[(\textstyle\sum_{p\in U_1} g(p)\delta_p')^2(\sum_{p\in U_2} g(p)\delta_p')^2] \\ \leq 17\left(\sum_{p\in U_1}\frac{g^2(p)}{p}\right)\left(\sum_{p\in U_2}\frac{g^2(p)}{p}\right).\end{aligned} \tag{4.16}$$

PROOF. The arguments are like those for Lemma 3.1. Let $h_i = \sum_{p\in U_i} g(p)\delta_p$, $\zeta_i = \sum_{p\in U_i} g(p)d_p$, $A_i = \sum_{p\in U_i} g(p)/p$, $B_i^2 = \sum_{p\in U_i} g^2(p)/p$, and $u_i = \max U_i$. Now

$$\mathbb{E}_n[h_1^{k_1}h_2^{k_2}] = \sum g(p_1)\cdots g(p_{k_1})g(q_1)\cdots g(q_{k_2})\mathbb{E}_n[\delta_{p_1}\cdots\delta_{p_{k_1}}\delta_{q_1}\cdots\delta_{q_{k_2}}]$$

where $p_1,\cdots,p_{k_1}$ range over U_1 and $q_1,\cdots,q_{k_2}$ range over U_2; and $\mathbb{E}[\zeta_1^{k_1}\zeta_2^{k_2}]$ has the analogous expansion. Just as we proved (3.15) by comparing (3.13) and (3.14), we can therefore prove

$$\begin{aligned}|\mathbb{E}_n[h_1^{k_1}h_2^{k_2}] - \mathbb{E}[\zeta_1^{k_1}\zeta_2^{k_2}]| &\leq \frac{1}{n}[\textstyle\sum_{p\in U_1}|g(p)|]^{k_1}[\sum_{p\in U_2}|g(p)|]^{k_2} \\ &\leq \frac{1}{n}(u_1B_1)^{k_1}(u_2B_2)^{k_2}\,.\end{aligned}$$

If in the argument leading to (3.16) we use a double binomial expansion, we arrive at

$$|\mathbb{E}_n[(h_1 - A_1)^{k_1}(h_2 - A_2)^{k_2}] - \mathbb{E}[(\zeta_1 - A_1)^{k_1}(\zeta_2 - A_2)^{k_2}]| \leqq \frac{1}{n}(|A_1| + u_1 B_1)^{k_1}(|A_2| + u_2 B_2)^{k_2} \leqq \frac{1}{n}(2u_1 B_1)^{k_1}(2u_2 B_2)^{k_2}.$$

If $k_1 = k_2 = 2$, the final term here is at most $16B_1^2B_2^2$ because $u_i \leqq n^{\frac{1}{8}}$. Since U_1 and U_2 are disjoint and the d_p are independent, $\mathbb{E}[(\zeta_1 - A)^2(\zeta_2 - A_2)^2] \leqq B_1^2B_2^2$, and (4.16) follows. □

Let $g(p)$ be defined over primes, as before, and for $u \leqq v$ let

$$M''_{uv} = \sup \min [|\textstyle\sum_{x_1<p\leqq x} g(p)\delta_p'|, |\sum_{x<p\leqq x_2} g(p)\delta_p'|],$$

where the supremum extends over triples x_1, x, x_2 with $u \leqq x_1 \leqq x \leqq x_2 \leqq v$.

LEMMA 4.2. *There is a universal constant C_1 such that, if $u \leqq v \leqq n^{\frac{1}{8}}$, then*

$$\mathbb{P}_n[M''_{uv} \geqq \lambda] \leqq \frac{C_1}{\lambda^4}\left[\sum_{u<p\leqq v} \frac{g^2(p)}{p}\right]^2. \tag{4.17}$$

PROOF. By Lemma 4.1,

$$\mathbb{E}_n[(\textstyle\sum_{x_1<p\leqq x} g(p)\delta_p')^2(\sum_{x<p\leqq x_2} g(p)\delta_p')^2] \leqq 17\left[\sum_{x_1<p\leqq x_2} \frac{g^2(p)}{p}\right]^2$$

for $x_1 \leqq x \leqq x_2 \leqq n^{\frac{1}{8}}$. Therefore (4.17) follows by Theorem 12.5 of [3] or Theorem 6.1 of [5]; the constant C_1, which may be taken as $17K''_{2,1}$ in the notation of [3] or as $17K$ in the notation of [5], is independent of n, u, v, g and λ. □

Because of (4.7), Lemma 4.2 applies to f_n. For $0 \leqq r \leqq s \leqq 1$, let

$$\textstyle\sum_{rs} = \sum_{p \in Q_n(s) - Q_n(r)} \frac{f_n^2(p)}{B_n^2 p};$$

if in (4.17) we take u to be the largest prime in $Q_n(r)$ and v to be the largest one in $Q_n(s)$, we obtain

$$\mathbb{P}_n[\sup \min [|X_n(t) - X_n(t_1)|, |X_n(t_2) - X_n(t)|] \geqq \varepsilon] \leqq \frac{C_1}{\varepsilon^4} \textstyle\sum_{rs}^2,$$

where the supremum extends over triples t_1, t, t_2 with $r \leqq t_1 \leqq t \leqq t_2 \leqq s$. It follows by Theorem 11.3 that, if

$$0 = s_0 < s_1 < \cdots < s_k = 1, \qquad s_i - s_{i-1} \geqq \delta, \qquad i = 1, \cdots, k, \tag{4.18}$$

then

$$\mathbb{P}_n[w''(X_n, \delta) > \varepsilon] \leqq \frac{C_1}{\varepsilon^4} \textstyle\sum_{i=0}^{k-2} \sum_{s_i, s_{i+2}}^2.$$

Clearly $\sum_{i=0}^{k-2} \sum_{s_i, s_{i+2}} \leqq 2$, and by (4.13), $\sum_{s_i, s_{i+2}} \leqq s_{i+2} - s_i + 2\varepsilon_n$ where ε_n is the maximum in (4.13) and (4.14). If we choose the s_i to satisfy $s_i - s_{i-1} \leqq 2\delta$ as well as (4.18), we obtain

$$\mathbb{P}_n[w''(X_n, \delta) > \varepsilon] \leqq \frac{2C_1}{\varepsilon^4}(4\delta + 2\varepsilon_n). \tag{4.19}$$

Given positive ε and η, choose δ so that $12C_1\delta/\varepsilon^4 < \eta$; by (4.14), $\varepsilon_n < \delta$ for large n, in which case the right side of (4.19) is less than η. Therefore, for each positive ε and η there exist a δ, $0 < \delta < 1$, and an integer n_0 such that

$$\mathbb{P}_n[w''(X_n, \delta) > \varepsilon] < \eta \,, \qquad n \geqq n_0 \,. \tag{4.20}$$

Now (4.6) follows by Theorem 15.4 of [3], which completes the proof of Theorem 4.1.

5. Approximation by infinitely divisible laws. This section extends to the general case of an infinitely divisible limit the central limit theorem of Section 3. The proof of that theorem required a sequence $\{r_n\}$ satisfying

$$\frac{\log r_n}{\log n} \to 0 \tag{5.1}$$

and

$$\frac{1}{B_n^2} \sum_{r_n < p \leqq n} \frac{f_n^2(p)}{p} \to 0 \tag{5.2}$$

(see (3.5) and (3.6)). Such a sequence in general need not exist, and to carry the theory through, we must restrict ourselves to those arrays for which one does exist. Let H be the class of arrays $\{f_n\}$ of additive functions for which (5.1) and (5.2) hold for some $\{r_n\}$ and for which the further condition

$$\lim_{n\to\infty} \frac{f_n(m)}{B_n} = 0 \,, \qquad m = 1, 2, \cdots \,, \tag{5.3}$$

holds. If f_n does not depend on n, then (5.3) is equivalent to $B_n \to \infty$, and the first condition is that $B_{r_n} \sim B_n$ for some $\{r_n\}$ satisfying (5.1).

If will be convenient in this section to omit the normalization by B_n; this can be arranged by passing from f_n to f_n/B_n. Let H' be the class of arrays $\{f_n\}$ of additive functions for which

$$\sum_{r_n < p \leqq n} \frac{f_n^2(p)}{p} \to 0 \tag{5.4}$$

for some sequence $\{r_n\}$ satisfying (5.1), for which

$$\sup_n \sum_{p \leqq n} \frac{f_n^2(p)}{p} < \infty \,, \tag{5.5}$$

and for which

$$\lim_{n\to\infty} f_n(m) = 0 \,, \qquad m = 0, 1, \cdots \,. \tag{5.6}$$

If $\{f_n\}$ is in H, then $\{f_n/B_n\}$ is in H'; not all arrays in H' arise in this way, since the sums in (5.5) need not equal 1.

Let K_n be the finite measure on the line R^1 corresponding to a mass of $f_n^2(p)/p$ at the point $f_n(p)$ for each $p \leqq n$:

$$K_n(M) = \sum_{p \leqq n, f_n(p) \in M} \frac{f_n^2(p)}{p} \,. \tag{5.7}$$

If K is a finite measure on the line, there is an infinitely divisible distribution F_K with characteristic function

(5.8) $$\varphi_K(u) = \exp \int_{-\infty}^{\infty} (e^{iux} - 1 - iux) \frac{1}{x^2} K(dx) ;$$

F_K has mean 0 and variance $K(R^1)$. The general infinitely divisible distribution with mean 0 and finite variance has this form. If K is a unit mass at 0, then $F_K = \Phi$.

Write $K_n \rightarrow_v K$ to indicate vague convergence: $K_n(M) \rightarrow K(M)$ for every finite interval M whose endpoints are continuity points of K. Weak convergence $K_n \Rightarrow K$ for finite measures can be defined by requiring $K_n \rightarrow_v K$ and $K_n(R^1) \rightarrow K(R^1)$.

THEOREM 5.1. *If $\{f_n\}$ is in H', then a necessary and sufficient condition for*

(5.9) $$f_n - \sum_{p \leqq n} \frac{f_n(p)}{p} \Rightarrow F$$

is that F have the form F_K and

(5.10) $$K_n \rightarrow_v K .$$

It is not difficult to see that the equivalence of these conditions persists if (5.10) is strengthened to $K_n \Rightarrow K$ and if to (5.9) is added the requirement that $\sum_{p \leqq n} f_n^2(p)/p$ converge to the variance of F.

PROOF OF SUFFICIENCY. It will be clear that the argument to follow still holds if n goes to infinity through some subsequence of the integers, a fact that will be needed in the proof of necessity.

Assume (5.10); we are to prove (5.9) with $F = F_K$. By (5.6) and Lemma 2.2, we may assume f_n to be completely additive; then

(5.11) $$\sum_{p \leqq n} f_n(p)\delta_p' \Rightarrow F_K$$

is the objective. The probability analogue is

(5.12) $$\sum_{p \leqq n} f_n(p)d_p' \Rightarrow F_K .$$

As the first step in the analysis of (5.12), we show that

(5.13) $$[f_n(p)d_p' : p \leqq n]$$

is an infinitesimal array (see (11.5)). We have

(5.14) $$\sum_{p \leqq n} \frac{f_n^2(p)}{p^2} \rightarrow 0 ;$$

indeed, if S^2 is the supremum in (5.5), then the sum here is at most $\sum_{p \leqq \varepsilon^{-1}} f_n^2(p) + \varepsilon S^2$, so that (5.14) follows from (5.6). Since (5.14) implies

(5.15) $$\max_{p \leqq n} \frac{|f_n(p)|}{p} \rightarrow 0 ,$$

it suffices to prove that $[f_n(p)d_p : p \leqq n]$ is infinitesimal, or that

$$\max_{p \leqq n} \mathbb{P}[|f_n(p)d_p| > \varepsilon] = \max\left[\frac{1}{p} : p \leqq n, |f_n(p)| > \varepsilon\right] \tag{5.16}$$

goes to 0. By (5.6), there exists an n_ε for which $n \geqq n_\varepsilon$ implies that $|f_n(p)| \leqq \varepsilon$ holds for all $p \leqq 1/\varepsilon$. But then $n \geqq n_\varepsilon$ implies that (5.16) is less than ε.

Since the array (5.13) is infinitesimal, the general convergence theory applies (see Theorem 11.4). Let K_n' be the measure having for each $p \leqq n$ a mass

$$\left[f_n(p)\left(0 - \frac{1}{p}\right)\right]^2\left(1 - \frac{1}{p}\right) = f_n^2(p)\frac{1}{p^2}\left(1 - \frac{1}{p}\right) \tag{5.17}$$

at the point $f_n(p)(0 - 1/p)$ and a mass

$$\left[f_n(p)\left(1 - \frac{1}{p}\right)\right]^2\frac{1}{p} = f_n^2(p)\frac{1}{p}\left(1 - \frac{1}{p}\right)^2 \tag{5.18}$$

at the point $f_n(p)(1 - 1/p)$. Since by (5.5) the row sums of (5.13) have bounded variances, (5.12) holds if

$$K_n' \to_v K\,. \tag{5.19}$$

The total weight of the masses (5.17) goes to 0 by (5.14), so removing them from K_n' has no effect on (5.19). If each mass of the form (5.18) is increased to $f_n^2(p)/p$, the total increase goes to 0 by (5.14), so (5.19) is again unchanged. Finally, if each mass of the form (5.18) is shifted from the point $f_n(p)(1 - 1/p)$ to the point $f_n(p)$, the maximum shift goes to 0 by (5.15), and (5.19) is once more unaffected. But with these three changes, K_n' becomes K_n, so that (5.19) becomes (5.10).

Therefore (5.10) implies (5.12). We shall show that (5.12) implies (5.11), but first under the added assumption

$$\sup_n \max_{p \leqq n} |f_n(p)| < \infty\,. \tag{5.20}$$

Because of (5.4), it follows by Lemma 2.1 that (5.11) is equivalent to

$$\textstyle\sum_{p \leqq r_n} f_n(p)\delta_p' \Rightarrow F_K\,. \tag{5.21}$$

By Chebyshev's inequality, (5.12) is equivalent to

$$\textstyle\sum_{p \leqq r_n} f_n(p)d_p' \Rightarrow F_K\,. \tag{5.22}$$

Thus (5.22) holds, and we are to deduce (5.21) from it. Let h_n and ζ_n be the nth terms in (5.21) and (5.22). By Lemma 3.1, $|\mathbb{E}_n[h_n^k] - \mathbb{E}[\zeta_n^k]| \leqq 2^k r_n^k S^k/n$, S^2 being the supremum in (5.5). If M exceeds the supremum in (5.20) and $M \geqq 1$, then $|\mathbb{E}[\zeta_n^k]| \leqq k!\, M^k e^{kS^2}$ by Lemma 3.2. Thus (5.22) implies (5.21) by Theorem 11.2.

To treat the general case, where (5.20) need not hold, define for each positive T a completely additive $f_{T,n}$ by

$$f_{T,n}(p) = f_n(p) \quad \text{if} \quad |f_n(p)| \leqq T$$
$$= 0 \quad \text{otherwise.}$$

Let $K_{T,n}$ be the measure with mass $f_{T,n}^2(p)/p$ at $f_{T,n}(p)$ for each $p \leqq n$, so that $K_{T,n}(M) = K_n(M \cap [-T, T])$. Define K_T by $K_T(M) = K(M \cap [-T, T])$. If $\pm T$ are continuity points of K, then (5.10) implies that $K_{T,n} \to_v K_T$ as $n \to \infty$. It is easy to see that $\{f_{T,n}\}$ lies in H' (the same $\{r_n\}$ works as for $\{f_n\}$). Since (5.20) holds with $f_{T,n}$ in place of f_n, it follows by the case already treated that

$$\sum_{p\leqq n} f_{T,n}(p)\delta_p' \Rightarrow F_{K_T} \qquad (n \to \infty) \tag{5.23}$$

if $\pm T$ are continuity points of K. Now $K_T \to_v K$ as $T \to \infty$, and hence $F_{K_T} \Rightarrow F_K$. By Theorem 11.1 (applied with T going to infinity through a sequence of points for which $\pm T$ are continuity points of K), (5.11) will follow from (5.23) if we show that

$$\lim_{T\to\infty} \limsup_{n\to\infty} \mathbb{P}_n[|\sum_{p\leqq n} (f_n(p) - f_{T,n}(p))\delta_p'| \geqq \varepsilon] = 0 . \tag{5.24}$$

If $K_n \Rightarrow K$, (5.24) follows easily from Lemma 2.1. The general case requires a further argument. The sum in (5.24) is

$$\begin{aligned}\sum f_n(p)\delta_p' &= \sum \frac{f_n(p)}{p}\delta_p' + \sum \frac{f_n(p)}{p}\left(\frac{1}{p} - 1\right) + \sum f_n(p)\left(1 - \frac{1}{p}\right)\delta_p \\ &= \sum\nolimits_1 + \sum\nolimits_2 + \sum\nolimits_3 ,\end{aligned}$$

where each sum here extends over those p with $p \leqq n$ and $|f_n(p)| > T$. By Lemma 2.1 and (5.14),

$$\mathbb{E}_n[\sum\nolimits_1^2] \leqq C \sum_{p\leqq n} \frac{f_n^2(p)}{p^3} \to 0 \qquad (n \to \infty) .$$

As for the nonrandom sum $\sum_2$,

$$|\sum\nolimits_2| \leqq \sum \frac{|f_n(p)|}{p} \leqq \frac{1}{T} \sum \frac{f_n^2(p)}{p} \leqq \frac{S^2}{T} ,$$

the sums having the same range as before and S^2 being the supremum in (5.5). Finally, if $T > 1$ then

$$\mathbb{P}_n[\sum\nolimits_3 \neq 0] \leqq \sum \frac{1}{p} \leqq \sum \frac{|f_n(p)|}{p} \leqq \frac{S^2}{T} ,$$

the sums with the same range. If $S^2/T < \varepsilon/2$, then the probability in (5.24) is at most $4\varepsilon^{-2}\mathbb{E}_n[\sum_1^2] + S^2/T \to S^2/T$ $(n \to \infty)$. This proves (5.24). □

PROOF OF NECESSITY. Since $\sup_n K_n(R^1) < \infty$ by (5.5), it follows from Helley's theorem that each subsequence $\{K_{n'}\}$ contains a further subsequence $\{K_{n''}\}$ such that

$$K_{n''} \to_v K \tag{5.25}$$

for some finite measure K. By the sufficiency part of the theorem, already established, $f_{n''} - \sum_{p\leqq n''} f_{n''}(p)/p \Rightarrow F_K$. But if (5.9) holds, then F_K must coincide with F. Since the K in the representation (5.8) is unique, there is thus only one possible limit in (5.25), which must therefore be the limit of the entire sequence $\{K_n\}$. □

We can now extend Theorem 3.1 to the general Lindeberg case and prove a partial converse.

THEOREM 5.2. *If* $\{f_n\}$ *is an array of additive functions satisfying* (5.6) *and*

$$\sum_{p\leqq n}\frac{f_n^2(p)}{p}\to 1\,, \tag{5.26}$$

then

$$\sum_{p\leqq n,|f_n(p)|>\varepsilon}\frac{f_n^2(p)}{p}\to 0\,, \qquad \varepsilon>0\,, \tag{5.27}$$

implies

$$f_n-\sum_{p\leqq n}\frac{f_n(p)}{p}\Rightarrow\Phi\,. \tag{5.28}$$

If $\{f_n\}$ *lies in* H' *and satisfies* (5.26), *then* (5.28) *implies* (5.27).

PROOF. Because of Theorem 5.1, we need only show that (5.6), (5.26), and (5.27) together imply that $\{f_n\}$ is in H'. Now (5.27) implies the existence of an increasing sequence $\{n_k\}$ such that $n\geqq n_k$ implies

$$\sum_{p\leqq n,|f_n(p)|>k^{-1}}\frac{f_n^2(p)}{p}<\frac{1}{k}\,.$$

If $r_n=n^{1/k}$ for $n_k\leqq n<n_{k+1}$, then certainly (5.1) holds. As for (5.4), split the sum according as $|f_n(p)|\leqq 1/k$ or not:

$$\sum_{r_n<p\leqq n}\frac{f_n^2(p)}{p}\leqq\frac{1}{k^2}\sum_{r_n<p\leqq n}\frac{1}{p}+\frac{1}{k}\,,\qquad n_k\leqq n<n_{k+1}\,.$$

Since by (1.21) the sum on the right is $O(\log k)$, (5.4) follows. □

Every F_K is a possible limit in Theorem 5.1. To see this, suppose first that K consists of positive masses $\mu_1,\cdots,\mu_l$ at points $x_1,\cdots,x_l$, where the x_i are distinct from each other and from 0. Choose $0<\theta_0<\theta_1<\cdots<\theta_l$ in such a way that $\log\theta_i/\theta_{i-1}=\mu_i/x_i^2$, $i=1,\cdots,l$, and define a completely additive f_n by

$$f_n(p)=x_i \quad \text{if} \quad (\log n)^{\theta_{i-1}}<p\leqq(\log n)^{\theta_i}\,, \qquad i=1,\cdots,l \tag{5.29}$$
$$=0 \quad \text{otherwise.}$$

Then $\{f_n\}$ is in H' and $K_n\Rightarrow K$.

For the general (finite) K, construct measures $K^{(k)}$ of the preceding kind with $L(K,K^{(k)})<1/k$, where L denotes Lévy distance ([17] page 39). For each k, there exists by the above argument an array $\{f_{k,n}\}$ in H' such that $K_n^{(k)}\Rightarrow K^{(k)}$ as $n\to\infty$, $K_n^{(k)}$ being the measure (5.7) corresponding to $f_{k,n}$. Let $\{r_{k,n}\}$ be the sequence in (5.1) and (5.4) for $\{f_{k,n}\}$. There exists an increasing sequence $\{n_k\}$ such that $n\geqq n_k$ implies

$$\frac{\log r_{k,n}}{\log n}<\frac{1}{k}\,,\qquad \sum_{r_{k,n}<p\leqq n}\frac{f_{k,n}^2(p)}{p}<\frac{1}{k}\,,\qquad \max_{p\leqq k}|f_{k,n}(p)|<\frac{1}{k}\,, \tag{5.30}$$

and $L(K^{(k)}, K_n^{(k)}) < 1/k$. If $f_n = f_{k,n}$ for $n_k \leqq n < n_{k+1}$, then $\{f_n\}$ is in H' and $K_n \Rightarrow K$. Thus every K and F_K are possible.

Suppose f_n is completely additive with

$$f_n(p) = 1 \quad \text{if} \quad r_n < p \leqq r_n^\theta \\ = 0 \quad \text{otherwise,} \tag{5.31}$$

where $\theta > 1$ and $\{r_n\}$ is any sequence going to infinity slowly enough that (5.1) holds. Then $\{f_n\}$ is in H', and the limit K consists of a mass of $\lambda = \log \theta$ at the point 1. Thus F_K is the distribution of $P_\lambda - \lambda$, where P_λ is a Poisson variable with mean λ. The centering sum in (5.9) can be added back in, since it converges to λ, and we may conclude that $f_n \Rightarrow P_\lambda$. Here $f_n(m)$ is the number of prime divisors of m in the range $r_n < p \leqq r_n^\theta$.

If f_n has the form f/B_n, the class of possible limits is much smaller, consisting of those K with density $|x|$ in some finite interval adjacent to 0 on the left, those K with a single mass at 0, those K with density x in some interval adjacent to 0 on the right, and linear combinations of these; see [22] page 64. This class of course includes the normal case and excludes the Poisson case. It is perhaps interesting to note that (if $f_n = f/B_n$) $\{f_n\}$ is in H if and only if B_n^2 goes to infinity in such a way that $B^2_{[e^x]}$ is a slowly varying function of x; hence the form of B_n^2 can be written out (see [15] **2** page 282).

6. Approximation by additive processes. We shall extend the functional limit theorem of Section 4, replacing Brownian motion by a more general additive process, or process of independent increments.

For $0 \leqq t \leqq 1$, let K_t be a measure on the line with

$$K_t(R^1) = t\,. \tag{6.1}$$

Suppose that, for each Borel set M,

$$K_s(M) \leqq K_t(M)\,, \qquad 0 \leqq s \leqq t \leqq 1\,, \tag{6.2}$$

so that $K_{s,t}(M) = K_t(M) - K_s(M)$ defines a measure with $K_{s,t}(R^1) = t - s$. By Theorem 15.7 of [3], the family $\{K_t\}$ determines a random element X of D, a random element whose increments are independent and have characteristic functions

$$\mathbb{E}[e^{iu(X(t)-X(s))}] = \exp \textstyle\int_{-\infty}^{\infty} (e^{iux} - 1 - iux)\, \dfrac{1}{x^2}\, K_{s,t}(dx)\,. \tag{6.3}$$

Since $X(t) - X(s)$ has variance $t - s$, X can have no fixed points of discontinuity. On the other hand, only in the case of Brownian motion will the paths of X be continuous with probability 1 (see Theorem 19.1 of [3]). For Brownian motion, K_t consists of a mass of t at 0.

The theory here goes through if (6.2) holds and $K_t(R^1)$ is continuous in t; the requirement (6.1) is only a convenient normalization.

For an additive array $\{f_n\}$, define $Q_n(t)$ by (4.1) and the function $X_n(t, m)$ by

(4.2)—or by (4.3) if f_n is completely additive. Let $K_{n,t}$ be the measure having a mass of $f_n^2(p)/B_n^2p$ at the point $f_n(p)/B_n$ for each p in $Q_n(t)$:

$$K_{n,t}(M) = \sum_{p \in Q_n(t), f_n(p)B_n^{-1} \in M} \frac{f_n^2(p)}{B_n^2 p} . \tag{6.4}$$

Then $K_{n,1}$ is the K_n of Section 5 for f_n/B_n—see (5.7).

THEOREM 6.1. *If* $\{f_n\}$ *is in* H *and if* $K_{n,t} \Rightarrow K_t$ *for each* t, *where* $\{K_t\}$ *satisfies* (6.1) *and* (6.2) *and hence determines an* X *via* (6.3), *then* $X_n \Rightarrow X$.

PROOF. Because of (5.3), one of the defining conditions of H, we may, just as in Section 4, restrict attention to completely additive f_n and take X_n to be defined by (4.3). And we may assume as before that (4.7) holds; to set $f_n(p) = 0$ for $p > n^{\frac{1}{2}}$ has asymptotically no effect on $K_{n,t}$ or, as shown in Section 4, on X_n. Now $K_{n,t}(R^1) \to t$ by hypothesis; since the functions are non-decreasing in t and the limit is continuous, the convergence must be uniform in t, and therefore the maximum jump in $K_{n,t}(R^1)$—as a function of t—goes to 0. Therefore (4.14) again holds.

The tightness argument in Section 4 (the proof of (4.20)) required of f_n only the properties (4.7) and (4.14), which hold here as well. Since X has no fixed points of discontinuity, Theorem 15.4 of [3] can again be applied. Therefore we need only prove that the finite-dimensional distributions converge.

We treat the two-dimensional case. Suppose $0 \leqq s \leqq t \leqq 1$, and define g_n by (4.10). We must prove

$$\frac{1}{B_n} \sum_{p \leqq n} g_n(p)\delta_p' \Rightarrow aX_s + bX_t .$$

Define V_n^2 by (4.12). We have seen that (4.14) holds again, and hence (4.15) follows just as before. It therefore suffices to prove that

$$\frac{1}{V_n} \sum_{p,n} g_n(p)\delta_p' \Rightarrow \theta(aX_s + bX_t) , \tag{6.5}$$

where $\theta = [(a+b)^2 s + b^2(t-s)]^{-\frac{1}{2}}$. From (4.15) and the assumption that $\{f_n\}$ is in H, it follows that $\{g_n/V_n\}$ is in H', and so Theorem 5.1 applies. Now the measure (5.7) for g_n/V_n is given by

$$K_n'(M) = \sum_{p \leqq n, g_n(p)V_n^{-1} \in M} \frac{g_n^2(p)}{V_n^2 p} .$$

Let ψ_z denote the function

$$\psi_z(x) = zx . \tag{6.6}$$

If $K_{n,s,t} = K_{n,t} - K_{n,s}$, and $\theta_n = B_n/V_n$, then

$$K_n' = \theta_n^2(a+b)^2 K_{n,s}\psi^{-1}_{\theta_n(a+b)} + \theta_n^2 b^2 K_{n,s,t}\psi^{-1}_{\theta_n b} .$$

Since X has independent increments, it follows by (6.3) and a change of variable that

$$\mathbb{E}[\exp[iu\theta(aX_s + bX_t)]] = \exp \int_{-\infty}^{\infty} (e^{iux} - 1 - iux) \frac{1}{x^2} K'(dx) ,$$

where K' is given by

$$K' = \theta^2(a + b)^2 K_s \psi_{\theta(a+b)}^{-1} + \theta^2 b^2 K_{s,t} \psi_{\theta b}^{-1} .$$

Since $K_{n,t} \Rightarrow K_t$, and since $\theta_n \to \theta$ by (4.15), it follows that $K_n' \Rightarrow K'$. Thus (6.5) is a consequence of Theorem 5.1. □

It is now a simple matter to prove Theorem 4.1 under the Lindeberg condition.

THEOREM 6.2. *If $\{f_n\}$ is an array of additive functions satisfying* (5.3) *and*

$$\text{(6.7)} \qquad \frac{1}{B_n^2} \sum_{p \leq n, |f_n(p)| > \varepsilon B_n} \frac{f_n^2(p)}{p} \to 0 , \qquad \varepsilon > 0 ,$$

then $X_n \Rightarrow W$.

PROOF. As in the proof of Theorem 5.2, from (5.3) and (6.7) it follows that $\{f_n\}$ is in H. Since (6.7) implies $K_{n,t}$ converges weakly to a mass of t at 0, and since these limits determine Brownian motion, $X_n \Rightarrow W$ follows by Theorem 6.1. □

Any family $\{K_t\}$ satisfying (6.1) and (6.2) can occur as a limit in Theorem 6.1. The proof involves the construction (5.29). Write $K' \dashv K''$ to indicate that $K'(M) \leq K''(M)$ for all Borel sets M. Suppose $0 = t_0 < t_1 < \cdots < t_u = 1$ and

$$\text{(6.8)} \qquad G_0 \dashv G_1 \dashv \cdots \dashv G_u , \qquad G_i(R^1) = t_i , \qquad i = 1, \cdots, u .$$

Suppose for the moment that each G_i has a finite support that excludes 0. For $1 \leq i \leq u$, construct completely additive functions $f_{i,n}$ by a formula like (5.29), in such a way that $\{f_{i,n}\}$ lies in H' and the corresponding measures (5.7) converge weakly as $n \to \infty$ to $G_i - G_{i-1}$. There is no difficulty in arranging that $\sum_p f_{i,n}^2(p)/p$ is exactly equal to $t_i - t_{i-1}$. If the θ_j in (5.29) are multiplied by a common constant, the values of the $\log \theta_j/\theta_{j-1}$ are unaltered and so the limiting distribution is also unaltered. Thus we may arrange that $f_{i,n}(p)$ vanishes unless $\alpha_{n,i-1} < p \leq \alpha_{n,i}$, where $0 < \alpha_{n,0} < \cdots < \alpha_{n,u}$.

Let $f_n = f_{1,n} + \cdots + f_{u,n}$, and let $J_{i,n}$ be the measure with a mass of $f_n^2(p)/p$ at $f_n(p)$ for each $p \leq \alpha_{n,i}$. By the construction, $B_n = 1$, $\{f_n\}$ lies both in H and in H', and $J_{i,n} \Rightarrow G_i$, $0 \leq i \leq u$, as $n \to \infty$. Moreover, $J_{i,n}(R^1)$ is exactly t_i, from which it follows that $[p : p \leq \alpha_{n,i}] \subset Q_n(t) \subset [p : p \leq \alpha_{n,j}]$ if $t_i \leq t < t_j$. Therefore the measures (6.4) for f_n satisfy

$$\text{(6.9)} \qquad J_{i,n} \dashv K_{n,t} \dashv J_{j,n} , \qquad t_i \leq t < t_j,\ 0 \leq i < j \leq u .$$

We have constructed $\{f_n\}$ in H and $\{J_{i,n}\}$ satisfying (6.9) and

$$\text{(6.10)} \qquad J_{i,n} \Rightarrow G_i \qquad (n \to \infty),\ 0 \leq i \leq u .$$

If the G_i are not concentrated on finite sets that exclude 0, the argument involving (5.30) can be applied (approximate the G_i by approximating the differences $G_i - G_{i-1}$). Thus (6.8) implies the existence of an array $\{f_n\}$ in H and measures $J_{i,n}$ for which (6.9) and (6.10) hold.

For each k, we now apply this result with $u = 2^k$, $t_i = i/2^k$, and $G_i = K_{i/2^k}$. Thus there exists an array $\{f_{k,n}\}$ in H and measures $J_{k,i,n}$ such that

$$J_{k,i,n} \prec K_{n,t}^{(k)} \prec J_{k,j,n}\,, \qquad \frac{i}{2^k} \leqq t < \frac{j}{2^k}\,, \qquad 0 \leqq i < j \leqq 2^k\,,$$

where $K_{n,t}^{(k)}$ is the measure (6.4) for $f_{k,n}$, and

$$J_{k,i,n} \Rightarrow K_{i/2^k} \qquad (n \to \infty),\ 0 \leqq i \leqq 2^k\,.$$

Choose an increasing sequence $\{n_k\}$ in such a way that $n \geqq n_k$ implies that (5.30) holds and the Lévy distance from $J_{k,i,n}$ to $K_{i/2^k}$ is less than $1/k$, $0 \leqq i \leqq 2^k$. If $f_n = f_{k,n}$ for $n_k \leqq n < n_{k+1}$, then $\{f_n\}$ is in H. Consider the measures $K_{n,t}$ corresponding to these f_n. Suppose $i/2^h \leqq t < j/2^h$. If $k > h$ and $n_k \leqq n < n_{k+1}$, then

$$J_{k,i2^{k-h},n} \prec K_{n,t} \prec J_{k,j2^{k-h},n}\,.$$

But the Lévy distance from the first of these measures to $K_{i2^{k-h}/2^k} = K_{i/2^h}$ is less than $1/k$, and similarly for the third measure. Thus, if t' and t'' are dyadic rationals with $t' \leqq t < t''$, there exist measures $J_{t',n}$ and $J_{t'',n}$ such that $J_{t',n} \prec K_{n,t} \prec J_{t'',n}$, $J_{t',n} \Rightarrow K_{t'}$, and $J_{t'',n} \Rightarrow K_{t''}$. Since $K_{t'} \prec K_t \prec K_{t''}$ and $K_{t''}(R^1) - K_{t'}(R^1) = t'' - t'$, a simple approximation argument shows that $K_{n,t} \Rightarrow K_t$. Thus any $\{K_t\}$ satisfying (6.1) and (6.2) is a possible limit in Theorem 6.1.

There is a functional limit theorem for (5.31). Suppose $\theta > 1$ and $\{r_n\}$ goes to infinity slowly enough that (5.1) holds. If, for $0 \leqq t \leqq 1$, $Y_n(t, m)$ is the number of prime divisors of m in the range $r_n < p \leqq r_n^{\theta t}$, then Y_n converges in distribution to a Poisson process with rate $\lambda = \log \theta$; for $\lambda = 1$ this is a consequence of Theorem 6.1, and so it clearly holds whatever λ is. It follows, for example, that the Lebesgue measure of $[t\colon Y_n(t, m) > t]$ has asymptotically a uniform distribution over $[0, 1]$ (see [21] page 261).

The hypotheses of Theorem 6.1 can be weakened in the case $f_n \equiv f$. Recall the definition (6.6).

THEOREM 6.3. *Suppose $\{f_n\}$ is in H and $f_n \equiv f$, and suppose*

$$\max_{p \leqq n} \frac{f^2(p)}{B_n^2 p} \to 0\,. \tag{6.11}$$

If $K_{n,1} \Rightarrow K$, then $K_{n,t} \Rightarrow K_t = tK\psi_{t^{1/2}}^{-1}$, where $\{K_t\}$ satisfies (6.1) and (6.2) and so determines an X to which X_n converges in distribution.

PROOF. If $v_{n,t}$ is the maximum element of $Q_n(t)$, then by (6.11) and the definition of $Q_n(t)$,

$$B^2_{v_{n,t}} = \sum_{p \in Q_n(t)} \frac{f^2(p)}{p} \sim tB_n^2\,. \tag{6.12}$$

Now

$$K_{n,t}(M) = \sum_{p \in Q_n(t),\, f(p)/B_n \in M} \frac{f^2(p)}{B_n^{\,2} p} = \frac{B^2_{v_{n,t}}}{B_n^{\,2}} K_{v_{n,t}1} \psi^{-1}_{B_{v_{n,t}}/B_n}(M) \,.$$

Now (6.12) and the hypothesis $K_{n,1} \Rightarrow K$ imply $K_{n,t} \Rightarrow K_t = tK\psi^{-1}_{t^{1/2}}$. Certainly $K_{n,t}(M)$ is non-decreasing in t, so (6.2) holds. Also, $K(R^1) = \lim_n K_{n,1}(R^1) = 1$, and (6.1) follows. □

The class of possible limits in Theorem 6.3 coincides with the class described at the end of Section 5; to prove this it is only necessary to observe that (6.11) is satisfied by the f constructed on page 65 of [22].

7. Sufficient conditions for a distribution. The results of this section and the next are analogues of the three-series theorem and hence are more naturally formulated for a single f than for an array. They give necessary and sufficient conditions for f to have a distribution in the sense of (1.30). See (1.10), (1.11) and (1.27) for the notation.

THEOREM 7.1. *If f is additive and the three series*

$$\sum_{|f(p)| \geqq c} \frac{1}{p}\,, \qquad \sum_{|f(p)| < c} \frac{f^2(p)}{p}\,, \qquad \sum_{|f(p)| < c} \frac{f(p)}{p} \tag{7.1}$$

converge for some $c > 0$, then

$$f \Rightarrow \sum_p f(p^{b_p})\,, \tag{7.2}$$

or, in case f is completely additive,

$$f \Rightarrow \sum_p f(p) d_p \,. \tag{7.3}$$

In the next section, we shall require a slightly different theorem.

THEOREM 7.2. *If the first two series in* (7.1) *converge, and if*

$$a_n = \sum_{|f(p)| < c,\, p \leqq n} \frac{f(p)}{p}\,, \tag{7.4}$$

then

$$f - a_n \Rightarrow \sum_{|f(p)| < c} \left[f(p^{b_p}) - \frac{f(p)}{p} \right] + \sum_{|f(p)| \geqq c} f(p^{b_p}) \,. \tag{7.5}$$

PROOFS. Suppose the first two of the series in (7.1) converge. If $f(p^{b_p}) \neq 0$, then $b_p \geqq 1$, which has probability $1/p$. Since the first series in (7.1) converges, the second series on the right in (7.5) converges with probability 1 by the Borel–Cantelli lemma. If $f(p^{b_p}) \neq f(p) d_p$, then $b_p \geqq 2$, which has probability $1/p^2$. Since $\sum 1/p^2$ converges, there is probability 1 that the first series on the right in (7.5) differs in only finitely many terms from $\sum_{|f(p)| < c} f(p) d_p'$. This last series converges with probability 1 by Kolmogorov's theorem ([9] page 108), because the middle series in (7.1) converges. Thus both series on the right in (7.5) converge with probability 1.

By the discrete form of Scheffé's theorem ([29] or [3] page 224), the convergence in (1.14) is preserved if each side is summed over an arbitrary collection of u-tuples $(k_1, \cdots, k_u)$. Applying this fact with $p_1, \cdots, p_u$ consisting of the primes not exceeding T, we see that

$$(7.6) \qquad \sum_{|f(p)|<c, p\leq T}\left[f(p^{\beta_p}) - \frac{f(p)}{p}\right] + \sum_{|f(p)|\geq c, p\leq T} f(p^{\beta_p}) \Rightarrow \sum_{|f(p)|<c, p\leq T}\left[f(p^{b_p}) - \frac{f(p)}{p}\right] + \sum_{|f(p)|\geq c, p\leq T} f(p^{b_p}) .$$

The right side of (7.6) converges in distribution as $T \to \infty$ to the right side of (7.5). Suppose we show that

$$(7.7) \qquad \lim_{T\to\infty} \limsup_{n\to\infty} \mathbb{P}_n\left[\left|\sum_{|f(p)|<c, T<p\leq n}\left[f(p^{\beta_p}) - \frac{f(p)}{p}\right]\right| \geq \varepsilon\right] = 0$$

and

$$(7.8) \qquad \lim_{T\to\infty} \limsup_{n\to\infty} \mathbb{P}_n[|\sum_{|f(p)|\geq c, T<p\leq n} f(p^{\beta_p})| \geq \varepsilon] = 0 .$$

It will then follow by Theorem 11.1 that

$$\sum_{|f(p)|<c, p\leq n}\left[f(p^{\beta_p}) - \frac{f(p)}{p}\right] + \sum_{|f(p)|\geq c, p\leq n} f(p^{\beta_p})$$

converges in distribution to the right side of (7.5), which is equivalent to (7.5).

Since $f(p^{\beta_p}) = f(p)\delta_p$ unless $\beta_p \geq 2$, Lemma 2.1 implies that the probability in (7.7) is at most

$$\sum_{p>T} \frac{1}{p^2} + \mathbb{P}_n[|\sum_{|f(p)|<c, T<p\leq n} f(p)\delta_p'| \geq \varepsilon] \leq \sum_{p>T} \frac{1}{p^2} + \frac{C}{\varepsilon^2} \sum_{|f(p)|<c, p>T} \frac{f^2(p)}{p} ,$$

which goes to 0 as $T \to \infty$. Since $f(p^{\beta_p}) = 0$ unless $\beta_p \geq 1$, the probability in (7.8) is at most

$$\sum_{|f(p)|\geq c, p>T} \frac{1}{p} ,$$

which also goes to 0 as $T \to \infty$. This proves (7.5).

If the third series in (7.1) also converges, say to a, then $a_n \to a$, and the convergence in (7.5) is preserved if we add a_n on the left and a on the right. But with this change, (7.5) reduces to (7.2). If f is completely additive, (7.2) obviously reduces to (7.3). □

By a theorem of Lévy [23], $\sum_p f(p^{b_p})$ has a continuous distribution function if

$$(7.9) \qquad \sum_{f(p)\neq 0} \frac{1}{p} = \infty ;$$

in the opposite case, the distribution is discrete by the Borel–Cantelli lemma.

Theorem 7.1 applies to the last three examples of Section 1.

EXAMPLE 3. Let $f(m) = \log(\sigma(m)/m)$. Since $\sigma(p^\beta) = 1 + p + \cdots + p^\beta$, $f(p) = \log(1 + 1/p)$ is of the order $1/p$, and the series (7.1) converge. Therefore

$$\log \frac{\sigma(m)}{m} \Rightarrow \sum_p \log\left(1 + \frac{1}{p} + \cdots + \frac{1}{p^{b_p}}\right).$$

Since (7.9) holds, the sum on the right has a continuous distribution function; therefore (see (1.29))

$$(7.10) \qquad D\left[m: \frac{\sigma(m)}{m} \leqq x\right] = \mathbb{P}\left[\Pi_p\left(1 + \frac{1}{p} + \cdots + \frac{1}{p^{b_p}}\right) \leqq x\right]$$

for all x. The integer m is said to be abundant if $\sigma(m) > 2m$, deficient if $\sigma(m) < 2m$, and perfect if $\sigma(m) = 2m$. By (7.10), the set of abundant numbers and the set of deficient numbers have densities, and the set of perfect numbers has density 0 (the last set is conjectured to be finite). The distribution function on the right in (7.10) is known [11] to be singular.

EXAMPLE 4. If $p_1, \cdots, p_k$ are the distinct prime factors of n, then Euler's function satisfies $\phi(n)/n = \mathbb{P}_n[m: p_i \nmid m, i = 1, \cdots, k]$. But since the p_i divide n, (1.6) implies that, under $\mathbb{P}_n$, the events $[m: p_i \mid m]$ have exact probabilities $1/p_i$, and (1.7) implies that they are independent (not merely approximately so). Therefore $\phi(n)/n = \Pi_i(1 - 1/p_i)$, which proves (1.5). If $f(m) = \log(\phi(m)/m)$, then f is completely additive and $f(p) = \log(1 - 1/p)$, so that series (7.1) again converge. Therefore

$$\log \frac{\phi(m)}{m} \Rightarrow \sum_p d_p \log\left(1 - \frac{1}{p}\right),$$

which implies

$$(7.11) \qquad \frac{\phi(m)}{m} \Rightarrow \Pi_p\left(1 - \frac{1}{p}\right)^{d_p}.$$

By (7.9), the distribution functions are again continuous.

Since the d_p are independent, the limiting variable in (7.11) has mean $\Pi_p(1 - p^{-2})$, which has the value ([19] page 245) $1/\zeta(2) = 6/\pi^2$. Since $\phi(m)/m$ is bounded, we can integrate the limit: $\mathbb{E}_n[\phi(m)/m] \to 6/\pi^2$. If $S_n = \sum_{m=1}^n \phi(m)/m$, then $S_n = n(6\pi^{-2} + \theta_n)$, where $\theta_n \to 0$. Now $\sum_{m=1}^n \phi(m) = nS_n - \sum_{m=1}^{n-1} S_m$, and so

$$\begin{aligned} \frac{1}{n^2} \sum_{m=1}^n \phi(m) &= \frac{6}{\pi^2} + \theta_n - \frac{1}{n^2} \sum_{m=1}^{n-1} m\left(\frac{6}{\pi^2} + \theta_m\right) \\ &= \frac{3}{\pi^2} + \theta_n + \frac{3}{n\pi^2} - \frac{1}{n^2} \sum_{m=1}^{n-1} m\theta_m . \end{aligned}$$

Since the last term here has modulus at most $n^{-1} \sum_{m=1}^n |\theta_m|$, which goes to 0, we have

$$\frac{1}{n^2} \sum_{m=1}^n \phi(m) \to \frac{3}{\pi^2} .$$

If two integers are chosen independently and randomly from among $1, 2, \cdots, n$, the chance they are relatively prime is $2n^{-2}\sum_{m=1}^{n}\phi(m)$, which is about $6/\pi^2$ for large n.

EXAMPLE 5. Let $f(m)$ be the excess $\Omega(m) - \omega(m)$. Since $f(p) = 0$, Theorem 7.1 applies, and

$$\Omega(m) - \omega(m) \Rightarrow \sum_p (b_p - d_p) .$$

As all these quantities are integer-valued,

$$D[m: \Omega(m) - \omega(m) = k] = \mathbb{P}[\sum_p (b_p - d_p) = k]$$

for nonnegative integers k. With $k = 0$, this gives

$$D[m: \Omega(m) = \omega(m)] = \Pi_p \left(1 - \frac{1}{p^2}\right) = \frac{6}{\pi^2} ; \tag{7.12}$$

$\Omega(m) = \omega(m)$ means that m is square-free—divisible by no perfect square.

8. Necessary conditions for a distribution. We prove the converse to Theorem 7.1.

THEOREM 8.1. *If the additive function f has a distribution, then the three series in* (7.1) *converge for every* $c > 0$.

The difficult part of the proof is to establish the existence of a λ for which

$$\sum_{|f(p)|\geq\lambda} \frac{1}{p} < \infty . \tag{8.1}$$

Let us assume (8.1), complete the proof of (7.1), and then return to (8.1).

We may assume $\lambda > c$. Since f has a distribution, for every ε there exists an x such that

$$\mathbb{P}_n[|f| \geqq x] < \varepsilon , \qquad n = 1, 2, \cdots , \tag{8.2}$$

(in other words, the sequence of distributions is tight). By (8.1) and by Theorem 7.1 with $c = \lambda$, $\sum_{|f(p)|\geqq\lambda} f(p^{\beta_p})$ has a distribution and hence for every ε there exists an x such that

$$\mathbb{P}_n[|\sum_{|f(p)|\geqq\lambda} f(p^{\beta_p})| \geqq x] < \varepsilon , \qquad n = 1, 2, \cdots .$$

It follows readily that the difference $f - \sum_{|f(p)|\geqq\lambda} f(p^{\beta_p})$ satisfies the same condition: for every ε there exists an x such that

$$\mathbb{P}_n[|\sum_{|f(p)|<\lambda} f(p^{\beta_p})| \geqq x] < \varepsilon , \qquad n = 1, 2, \cdots . \tag{8.3}$$

Define

$$u_n = \sum_{|f(p)|<\lambda, p\leqq n} \frac{f(p)}{p} , \qquad v_n^2 = \sum_{|f(p)|<\lambda, p\leqq n} \frac{f^2(p)}{p} . \tag{8.4}$$

If $v_n \to \infty$, then Theorem 3.1 implies that

$$\left|\mathbb{P}_n\left[\frac{1}{v_n} \sum_{|f(p)|<\lambda} f(p^{\beta_p}) - \frac{u_n}{v_n} \leqq x\right] - \Phi(x)\right| \leqq \varepsilon_n \to 0 .$$

But then the probability complementary to that in (8.3) is within $2\varepsilon_n$ of

$$\Phi\left(\frac{x-u_n}{v_n}\right)-\Phi\left(\frac{-x-u_n}{v_n}\right)\leqq\frac{1}{(2\pi)^{\frac{1}{2}}}\frac{2x}{v_n}\to 0\,,$$

contrary to (8.3). Therefore v_n is bounded. From this and (8.1), it follows that the first two of the series in (7.1) converge.

Define a_n by (7.4). According to Theorem 7.2, $f-a_n$ converges in distribution. If f has a distribution, then a_n must have a finite limit ([17] page 42), and so the third series in (7.1) must also converge.

To find a λ satisfying (8.1), we need two lemmas.

LEMMA 8.1. *Let f be additive, let Q be a set of primes p for which $f(p)>\lambda$, and let L be a set of square-free integers l for which $|f(l)|\leqq\lambda/2$. There is a universal constant C_0 such that, if*

$$\textstyle\sum_{p\in Q,p\leqq n}\frac{1}{p}=\tau\,,$$

then

$$\textstyle\sum_{l\in L,l\leqq n}\frac{1}{l}\leqq 1+C_0\left[\frac{1}{\tau^{\frac{1}{2}}}+\left(\frac{2}{e^{\frac{3}{4}}}\right)^{\tau}\right]\log n\,. \tag{8.5}$$

Note that $2<e^{\frac{3}{4}}$; C_0 is independent of f, n, λ, Q, and L.

PROOF. For any set M of integers, let $\Delta_M(m)$ be the number of divisors of m in M. Let Q' be the set of primes not in Q, and let S be the set of square-free integers composed of primes in Q'.

Clearly

$$\textstyle\mathrm{E}_n[\Delta_L]=\sum_{l\in L}\mathrm{P}_n[m:l\,|\,m]\geqq\sum_{l\in L,l\leqq n}\frac{1}{l}-1 \tag{8.6}$$

and

$$\textstyle\mathrm{E}_n[\Delta_L]=\frac{1}{n}\sum_{m\leqq n,\Delta_Q(m)<\tau}\Delta_L(m)+\frac{1}{n}\sum_{m\leqq n,\Delta_Q(m)\geqq\tau}\Delta_L(m)\,. \tag{8.7}$$

Since each integer in L is square-free,

$$\Delta_L(m)\leqq 2^{\omega(m)}=2^{\Delta_Q(m)}2^{\Delta_{Q'}(m)}=2^{\Delta_Q(m)}\Delta_S(m)\,,$$

and so

$$\textstyle\frac{1}{n}\sum_{m\leqq n,\Delta_Q(m)<\tau}\Delta_L(m)\leqq 2^{\tau}\mathrm{E}_n[\Delta_S]\leqq 2^{\tau}\sum_{s\in S,s\leqq n}\frac{1}{s}\leqq 2^{\tau}\Pi_{p\in Q',p\leqq n}\left(1+\frac{1}{p}\right).$$

By (1.21), there is a constant C_0 such that

$$\Pi_{p\leqq n}\left(1+\frac{1}{p}\right)\leqq C_0\log n\,. \tag{8.8}$$

Since

$$\Pi_{p\in Q,p\leqq n}\left(1+\frac{1}{p}\right)\geqq\Pi_{p\in Q,p\leqq n}\exp\frac{3}{4p}\,,$$

we have

$$\frac{1}{n} \sum_{m \leq n, \Delta_Q(m) < \tau} \Delta_L(m) \leqq C_0 2^\tau \frac{\log n}{e^{3\tau/4}} . \tag{8.9}$$

Fix m for the moment and suppose that $\Delta_Q(m) = x$. Then m has the form $m = rst$, where $r = q_1 \cdots q_x$, the q_i being distinct elements of Q, where s lies in S, and where t is composed of primes already appearing in rs (rs is the square-free part of m). If $l \mid m$ and $l \in L$, then, l being square-free, $l = r's'$ with $r' \mid r$ and $s' \mid s$. Now distinct values of r' with the same value of s' cannot divide: Suppose $l_1 = r_1's'$ and $l_2 = r_2's'$ ($r_1' \neq r_2'$), with $r_1' \mid r$, $r_2' \mid r$, and $s' \mid s$. Then $r_1' \nmid r_2'$ because $r_1' \mid r_2'$ would imply $l_1 \mid l_2$ and (by the properties of L and Q) $\lambda \geqq f(l_2) - f(l_1) = f(r_2') - f(r_1') = f(r_2'/r_1') > \lambda$, an impossibility. A collection of divisors of $q_1 \cdots q_x$, with no divisor dividing another one, has cardinality at most γ_x, where by (11.6), $\gamma_x \leqq C_0 2^x / x^{\frac{1}{2}}$ if C_0 is increased sufficiently. If G is the set of all square-free integers, then

$$\Delta_L(m) \leqq \gamma_x \Delta_S(m) \leqq C_0 \frac{2^x}{x^{\frac{1}{2}}} \Delta_S(m) = \frac{C_0}{x^{\frac{1}{2}}} \Delta_G(m) .$$

This holds for any m such that $\Delta_Q(m) = x$. By (8.8),

$$\frac{1}{n} \sum_{m \leq n, \Delta_Q(m) \geqq \tau} \Delta_L(M) \leqq \frac{C_0}{\tau^{\frac{1}{2}}} \mathbb{E}_n[\Delta_G] \leqq \frac{C_0}{\tau^{\frac{1}{2}}} \sum_{g \in G, g \leq n} \frac{1}{g}$$
$$\leqq \frac{C_0}{\tau^{\frac{1}{2}}} \Pi_{p \leq n} \left(1 + \frac{1}{p}\right) \leqq \frac{C_0^2}{\tau^{\frac{1}{2}}} \log n .$$

Putting this and (8.9) into (8.7) and using (8.6) gives (8.5) with a new C_0. □

LEMMA 8.2. *If*

$$\liminf_{n \to \infty} \mathbb{P}_n(L) > 0 , \tag{8.10}$$

then

$$\liminf_{n \to \infty} \frac{1}{\log n} \sum_{m \in L, m \leq n} \frac{1}{m} > 0 . \tag{8.11}$$

PROOF. Choose an integer θ so that $\mathbb{P}_n(L) > 2/\theta$ for large n. Then

$$\sum_{m \in L, k < m \leq \theta k} \frac{1}{m} \geqq \mathbb{P}_{\theta k}(L) - \frac{1}{\theta} \mathbb{P}_k(L) \geqq \frac{1}{\theta}$$

for large k. Given n, choose i so that $\theta^i \leqq n < \theta^{i+1}$; then sum $1/m$ over the blocks $\theta^j \leqq m < \theta^{j+1}$. □

We can now complete the proof of Theorem 8.1 by showing that, if for each ε there is an x satisfying (8.2), then (8.1) holds for some λ. By symmetry it is enough to find a λ such that

$$\sum_{f(p) > \lambda} \frac{1}{p} < \infty . \tag{8.12}$$

By (8.2) and (7.12), there is a λ such that, if L is the set of square-free integers

l with $|f(l)| \leqq \lambda/2$, then (8.10) holds and hence so does (8.11). Apply Lemma 8.1 with this L and with Q the set of p with $f(p) > \lambda$: if (8.12) fails, then (8.11) and (8.5) are in contradiction. This proves (8.1).

9. Entropy. As observed in Section 1, it takes $\log n$ units of information to specify an integer in the range $1, 2, \cdots, n$ and hence $\log n$ units of information to specify its factorization into primes. Since $\log n$ is unbounded, it will follow that there must be infinitely many primes if it is shown that each prime contributes to the factorization at most, say, 2 units of information.

A finite partition $\mathscr{A}$ of a probability space has entropy $H(\mathscr{A}) = -\sum P(A) \times \log P(A)$, the sum extending over the elements of the partition; see Chapter 2 of [2]. The common refinement $\bigvee_i \mathscr{A}_i$ of partitions $\mathscr{A}_1, \cdots, \mathscr{A}_r$ has as its elements the various intersections $A_1 \cap \cdots \cap A_r$ with A_i in $\mathscr{A}_i$. A basic fact of information theory is that entropy is subadditive in the sense that $H(\bigvee_i \mathscr{A}_i) \leqq \sum_i H(\mathscr{A}_i)$.

It will be convenient in this section to work in the space $\Omega_n = [1, 2, \cdots, n]$ and to take $\mathbb{P}_n$ as the measure on Ω_n assigning probability $1/n$ to each point. For prime p, let $\mathscr{A}_p{}^n$ be the partition of Ω_n with elements

$$(9.1) \qquad A_p{}^n(k) = [m \in \Omega_n : p^k \mid m, p^{k+1} \nmid m]$$

(the set is empty if $p^k > n$). To know which element of $\mathscr{A}_p{}^n$ an integer m lies in (m in the range $1 \leqq m \leqq n$) is to know the exact power of p in the factorization of m. By the fundamental theorem of arithmetic, $\bigvee_{p \leqq n} \mathscr{A}_p{}^n$ has the individual points of Ω_n as its elements and hence has entropy $\log n$; subadditivity now implies

$$(9.2) \qquad \log n = H(\bigvee_{p \leqq n} \mathscr{A}_p{}^n) \leqq \sum_{p \leqq n} H(\mathscr{A}_p{}^n) .$$

To estimate $H(\mathscr{A}_p{}^n)$, observe that

$$(9.3) \qquad \mathbb{P}_n(A_p{}^n(k)) = \frac{1}{n}\left[\frac{n}{p^k}\right] - \frac{1}{n}\left[\frac{n}{p^{k+1}}\right] \leqq \frac{1}{p^k} \leqq \frac{1}{2^k} .$$

Write $h(t) = -t \log t$; h is bounded by e^{-1} and is increasing for $t < e^{-1}$. Since $2^{-k} < e^{-1}$ for $k \geqq 2$, (9.3) gives

$$(9.4) \qquad H(\mathscr{A}_p{}^n) \leqq \frac{2}{e} + \sum_{k=2}^{\infty} h\left(\frac{1}{2^k}\right) = \frac{2}{e} + \tfrac{3}{2} \log 2 < 2 .$$

By (9.2) and (9.4), $\pi(n) \geqq \frac{1}{2} \log n$, and there are indeed infinitely many primes.

These ideas can be carried further. By (9.3),

$$(9.5) \qquad \sum_{k=2}^{\infty} h(\mathbb{P}_n(A_p{}^n(k))) \leqq \sum_{k=2}^{\infty} h\left(\frac{1}{p^k}\right) \leqq \frac{\log p}{p^2} \sum_{k=2}^{\infty} \frac{k}{2^{k-2}} = 6\frac{\log p}{p^2} .$$

Since h is concave and has slope -1 at $t = 1$, $h(t) \leqq 1 - t$ for all t and hence $h(\mathbb{P}_n(A_p{}^n(0))) = h(1 - n^{-1}[n/p]) \leqq 1/p$. On the other hand, $\log t \leqq t - 1$ for all t, so that $h(t) \geqq t(1-t)$ and hence $h(\mathbb{P}_n(A_p^n(0))) \geqq n^{-1}[n/p](1-n^{-1}[n/p]) \geqq p^{-1}-p^{-2}-n^{-1}$. Therefore

$$\left|h(\mathbb{P}_n(A_p^n)0))) - \frac{1}{p}\right| \leqq \frac{1}{p^2}+\frac{1}{n}. \tag{9.6}$$

Further,

$$h(\mathbb{P}_n(A_p^n(1))) = h\left(\frac{1}{n}\left[\frac{n}{p}\right]-\frac{1}{n}\left[\frac{n}{p^2}\right]\right) \geqq \left(\frac{1}{p}-\frac{1}{p^2}-\frac{1}{n}\right)\log p.$$

If $p \geqq 3$, then, since $p^{-1} \leqq e^{-1}$, we also have $h(\mathbb{P}_n(A_p^n(1))) \leqq h(p^{-1}) = p^{-1}\log p$. Therefore

$$\left|h(\mathbb{P}_n(A_p^n(1))) - \frac{\log p}{p}\right| \leqq \left(\frac{1}{p^2}+\frac{1}{n}\right)\log p, \qquad p \geqq 3. \tag{9.7}$$

By (9.5), (9.6) and (9.7),

$$\left|H(\mathscr{A}_p^n) - \frac{1}{p} - \frac{\log p}{p}\right| \leqq 2\left(\frac{1}{p^2}+\frac{1}{n}\right)\log p + 6\,\frac{\log p}{p^2}, \qquad p \geqq 3.$$

By (10.3), the right side here summed over $p \leqq n$ is $O(1)$. By (10.5) and (10.7) it therefore follows that

$$0 \leqq \textstyle\sum_{p\leqq n} H(\mathscr{A}_p^n) - H(\bigvee_{p\leqq n}\mathscr{A}_p^n) = \log\log n + O(1). \tag{9.8}$$

If the partitions $\mathscr{A}_p^n$ were independent under $\mathbb{P}_n$ (that is, if the β_p were independent random variables), the difference in (9.8) would vanish. That the difference is of smaller order than $H(\bigvee_{p\leqq n}\mathscr{A}_p^n) = \log n$ is an indication that the total amount of dependence is small, which is the fact exploited in the preceding sections.

The number-theoretic analogue of the conditional entropy (1.22) can be estimated in an elementary way. Let $\mathscr{B}_p^n$ be the partition of Ω_n with the two elements $[m \in \Omega_n : p \nmid m] = A_p^n(0)$ and $[m \in \Omega_n : p \mid m] = \Omega_n - A_p^n(0)$. To know which element of $\mathscr{B}_p^n$ contains m $(1 \leqq m \leqq n)$ is to know whether or not p divides m. Since $\mathscr{A}_p^n$ refines $\mathscr{B}_p^n$ (see [2] for the definition and properties of conditional entropy),

$$\begin{aligned} H(\mathscr{A}_p^n \mid \mathscr{B}_p^n) &= H(\mathscr{A}_p^n) - H(\mathscr{B}_p^n) \\ &= h\left(\frac{1}{n}\left[\frac{n}{p}\right]-\frac{1}{n}\left[\frac{n}{p^2}\right]\right) - h\left(\frac{1}{n}\left[\frac{n}{p}\right]\right) + \textstyle\sum_{k=2}^{\infty} h(\mathbb{P}_n(A_p^n(k))). \end{aligned}$$

If $p \geqq 3$, so that $p^{-1} < e^{-1}$, the second term on the right here exceeds the first, and (9.5) gives $H(\mathscr{A}_p^n \mid \mathscr{B}_p^n) < 6p^{-2}\log p$. Since $H(\mathscr{A}_2^n \mid \mathscr{B}_2^n) \leqq H(\mathscr{A}_2^n) < 2$, it follows that there exists a constant K such that $\sum_p H(\mathscr{A}_p^n \mid \mathscr{B}_p^n) < K$ for all n. Moreover, $H(\bigvee_p \mathscr{A}_p^n \mid \bigvee_q \mathscr{B}_q^n) \leqq \sum_p H(\mathscr{A}_p^n \mid \bigvee_q \mathscr{B}_q^n) \leqq \sum_p H(\mathscr{A}_p^n \mid \mathscr{B}_p^n)$, and hence

$$H(\textstyle\bigvee_p \mathscr{A}_p^n \mid \bigvee_p \mathscr{B}_p^n) < K \tag{9.9}$$

for all n.

A slight improvement of these estimates, together with some computation, shows that K may be taken as 2.6. Measured in bits rather than natural units, this is $2.6/\log 2$, which is less than 3.8.

10. Appendix: Number theory. By the definition (1.1) of $\beta_p(m)$, $\log m = \sum_p \beta_p(m) \log p$; let $\log^* m = \sum_p \delta_p(m) \log p$ (compare (2.15)). By Stirling's formula,

$$\mathbb{E}_n[\log] = \frac{1}{n} \log n! = \log n + O(1) . \tag{10.1}$$

Now $\mathbb{E}_n[\beta_p - \delta_p] = \sum_{k\geq 1} \mathbb{P}_n[\beta_p - \delta_p \geq k] \leqq \sum_{k\geq 1} 1/p^{k+1} \leqq 2/p^2$, and hence $\mathbb{E}_n[\log - \log^*] \leqq 2 \sum_p p^{-2} \log p$. By this and (10.1),

$$\mathbb{E}_n[\log^*] = \sum_p \frac{1}{n}\left[\frac{n}{p}\right] \log p = \log n + O(1) . \tag{10.2}$$

If $n < p \leqq 2n$, then $1 \leqq 2n/p < 2$ and $0 \leqq n/p < 1$, so that $[2n/p] - 2[n/p] = 1$. Therefore

$$\sum_{n<p\leqq 2n} \log p = 2n \sum_{n<p\leqq 2n} \left(\frac{1}{2n}\left[\frac{2n}{p}\right] - \frac{1}{n}\left[\frac{n}{p}\right]\right) \log p .$$

Since $2[y] \leqq [2y]$ (as $2[y] \leqq 2y$), the sum on the right does not decrease if extended over all p. The sum thus extended is $\mathbb{E}_{2n}[\log^*] - \mathbb{E}_n[\log^*]$, which by (10.2) is bounded. Thus there is a K such that $\sum_{n<p\leqq 2n} \log p \leqq Kn$ for all n.

Given $x > 1$, choose n so that $2^{n-1} \leqq x < 2^n$. Since

$$\sum_{p\leqq x} \log p \leqq \sum_{k=1}^{n} \sum_{2^{k-1}<p\leqq 2^k} \log p \leqq K \sum_{k=1}^{n} 2^{k-1} \leqq 2Kx ,$$

we have

$$\sum_{p\leqq x} \log p = O(x) . \tag{10.3}$$

If $\pi(x)$ is the number of primes not exceeding x, then

$$\pi(x) \leqq \sum_{p\leqq x^{1/2}} 1 + \sum_{x^{1/2}<p\leqq x} \frac{\log p}{\log x^{1/2}} \leqq x^{1/2} + \frac{2}{\log x} \sum_{p\leqq x} \log p ;$$

now (10.3) implies

$$\pi(x) = O\left(\frac{x}{\log x}\right) , \tag{10.4}$$

which is (1.20).

Removing the integral-part brackets in the middle term of (10.2) converts it into $\sum_{p\leqq n} p^{-1} \log p$ and introduces an error of at most $n^{-1} \sum_{p\leqq n} \log p$, which is $O(1)$ by (10.3). Hence, by (10.2), $\sum_{p\leqq n} p^{-1} \log p = \log n + O(1)$, and since $\log x - \log [x]$ is bounded,

$$\sum_{p\leqq x} \frac{\log p}{p} = \log x + O(1) , \tag{10.5}$$

which is (1.19).

According to the integration-by-parts formula, if dF and dG are mass distributions over $(a, b]$ with no common points of positive mass, then

$$\int_{(a,b]} F\,dG = FG\Big|_a^b - \int_{(a,b]} G\,dF\,. \tag{10.6}$$

Let dG have mass $p^{-1}\log p$ at each p, so that $G(t) = \sum_{p\leq t} p^{-1}\log p = \log t + \varphi(t)$, where φ is by (10.5) a bounded function. Let $F(t) = -1/\log t$, so that $dF(t) = dt/t\log^2 t$. Taking $a = \frac{3}{2}$ and $b = x$ in (10.6) gives

$$\begin{aligned}\sum_{p\leq x}\frac{1}{p} &= \frac{G(t)}{\log t}\Big|_a^x + \int_a^x \frac{\log t + \varphi(t)}{t\log^2 t}\,dt \\ &= 1 + \frac{\varphi(x)}{\log x} + \log\log x - \log\log a + \int_a^x \frac{\varphi(t)}{t\log^2 t}\,dt\,.\end{aligned}$$

Since $\varphi(t)$ is bounded, the last integral extended over (a, ∞) is finite and extended over (x, ∞) is $O(1/\log x)$. Therefore there is a constant c such that

$$\sum_{p\leq x}\frac{1}{p} = \log\log x + c + O\left(\frac{1}{\log x}\right), \tag{10.7}$$

which is (1.21).

Two other standard results, not needed in the paper, are easily obtained. If M bounds the $O(1)$ term in (10.5), then, for $0 < \theta < 1$ and all x, $\sum_{p\leq x}\log p \geqq \theta x \sum_{\theta x < p \leq x} p^{-1}\log p \geqq \theta x(\log\theta^{-1} - 2M)$. For small enough θ, the final factor here is positive, which, together with (10.3), implies

$$\sum_{p\leq x}\log p \asymp x$$

(the ratio of the two sides is bounded away from 0 and infinity). Finally, $\pi(x) \geqq \sum_{p\leq x}\log p/\log x \asymp x/\log x$, which, with (10.4), gives

$$\pi(x) \asymp \frac{x}{\log x}\,.$$

11. Appendix: Probability.

See Section 1 for notation. Suppose that $(\xi_n, \eta_{n,T})$ is a random vector for each n and T.

THEOREM 11.1. *If $\eta_{n,T} \Rightarrow \eta_T$ $(n \to \infty)$ for each T, if $\eta_T \Rightarrow \eta$ $(T \to \infty)$, and if*

$$\lim_{T\to\infty}\limsup_{n\to\infty} \mathbb{P}[|\xi_n - \eta_{n,T}| \geqq \varepsilon] = 0 \tag{11.1}$$

for $\varepsilon > 0$, then $\xi_n \Rightarrow \eta$.

PROOF. Let $\delta(n, T, \varepsilon)$ be the probability in (11.1). Suppose $x' < x < x''$, and choose ε smaller than $x'' - x$ and $x - x'$; then

$$\mathbb{P}[\eta_{n,T} \leqq x'] - \delta(n, T, \varepsilon) \leqq \mathbb{P}[\xi_n \leqq x] \leqq \mathbb{P}[\eta_{n,T} \leqq x''] + \delta(n, T, \varepsilon)\,.$$

If x' and x'' are continuity points of the distribution functions of η and all the η_T, it follows, upon letting $n \to \infty$ and then $T \to \infty$, that $\mathbb{P}[\xi_n \leqq x]$ has limits inferior and superior lying between $\mathbb{P}[\eta \leqq x']$ and $\mathbb{P}[\eta \leqq x'']$. These last two probabilities converge to $\mathbb{P}[\eta \leqq x]$ as $x', x'' \to x$, provided x is a continuity point of the distribution function of η. □

The above result can be found in more general form in [3] page 25.

THEOREM 11.2. *If* $\lim_n |\mathbb{E}[\xi_n{}^k] - \mathbb{E}[\zeta_n{}^k]| = 0$ *for* $k = 1, 2, \cdots$, *if there exists a* θ *such that* $|\mathbb{E}[\zeta_n{}^k]| \leqq k!\, \theta^k$ *for* $n, k = 1, 2, \cdots$, *and if* $\zeta_n \Rightarrow \zeta$, *then* $\xi_n \Rightarrow \zeta$.

PROOF. The hypotheses imply (see ([7] page 88 or [3] page 32) that $\lim_n \mathbb{E}[\zeta_n{}^k] = \mathbb{E}[\zeta^k]$. It follows that $|\mathbb{E}[\zeta^k]| \leqq k!\, \theta^k$, so that $\sum_k \mathbb{E}[\zeta^k] z^k/k!$ has positive radius of convergence and hence (see [15] **2** page 228 or [8] page 176) the distribution of ζ is determined by its moments. Now $\xi_n \Rightarrow \zeta$ follows because $\lim_n \mathbb{E}[\xi_n{}^k] = \mathbb{E}[\zeta^k]$ (see [7] page 92). □

For an element x of the space D, define

$$w''(x, \delta) = \sup \min [|x(t) - x(t_1)|, |x(t_2) - x(t)|] ,$$

where the supremeum extends over triples t_1, t, t_2 with $0 \leqq t_1 \leqq t \leqq t_2 \leqq 1$ and $t_2 - t_1 \leqq \delta$ (see [3] page 118).

THEOREM 11.3. *Let* X *be a random element of* D; *for* $0 \leqq r \leqq s \leqq 1$, *put*

$$\gamma(r, s) = \mathbb{P}[\sup \min [|X(t) - X(t_1)|, |X(t_2) - X(t)|] > \varepsilon] , \tag{11.2}$$

where the supremeum extends over triples with $r \leqq t_1 \leqq t \leqq t_2 \leqq s$. *If* $0 = s_0 < s_1 < \cdots < s_k = 1$ *and* $s_i - s_{i-1} \geqq \delta$, $i = 1, \cdots, k$, *then*

$$\mathbb{P}[w''(X, \delta) > \varepsilon] \leqq \textstyle\sum_{i=0}^{k-2} \gamma(s_i, s_{i+2}) . \tag{11.3}$$

PROOF. If $0 \leqq t_2 - t_1 \leqq \delta$, then $s_i \leqq t_1 \leqq t_2 \leqq s_{i+2}$ for some i, $0 \leqq i \leqq k - 2$, so (11.3) follows from the definition (11.2). □

Consider a triangular array: Suppose that, for each n, $\xi_{n1}, \cdots, \xi_{nk_n}$ are independent random variables. Suppose that

$$\mathbb{E}[\xi_{nk}] = 0 , \qquad \mathbb{E}[\xi_{nk}^2] = \sigma_{nk}^2 < \infty , \tag{11.4}$$

and suppose that the array is infinitesimal in the sense that

$$\max_{k \leq k_n} \mathbb{P}[|\xi_{nk}| \geqq \varepsilon] \to 0 \tag{11.5}$$

for $\varepsilon > 0$. Let K_n be the finite measure defined by

$$K_n(M) = \textstyle\sum_{k \leq k_n} \int_{\xi_{nk} \in M} \xi_{nk}^2 \, d\mathbb{P} ,$$

and define F_K by (5.8).

THEOREM 11.4. *If* (11.4) *and* (11.5) *hold, if* $\sum_{k \leq k_n} \sigma_{nk}^2 = O(1)$, *and if* $K_n \to_v K$, *then* $\sum_{k \leq k_n} \xi_{nk} \Rightarrow F_K$.

PROOF. This result is to be found on page 294 of [24], but with the assumption (11.5) strengthened to $\max_k \sigma_{nk}^2 \to 0$; however, this condition is used only in the "comparison lemma" on page 291, for which (11.5) itself actually suffices—see Theorem 1, page 98 of [17]. On the other hand, the result is to be found on page 100 of [17], but with the assumption $K_n \rightarrow_v K$ strengthened to $K_n \Rightarrow K$; this proof too is easily adapted to the present circumstance. □

We need a combinatorial result due to Sperner [30]. Denote the cardinality of a finite set by bars.

THEOREM 11.5. *Suppose $|S| = n$, and consider a class $\mathscr{A}$ of subsets of S with the property that no element of $\mathscr{A}$ is a subset of another element of $\mathscr{A}$. The maximum γ_n of $|\mathscr{A}|$ for such a class $\mathscr{A}$ satisfies*

$$\gamma_n = O\left(\frac{2^n}{n^{\frac{1}{2}}}\right). \tag{11.6}$$

PROOF. Suppose $\mathscr{A}$ has the required property and $|\mathscr{A}| = \gamma_n$. Suppose $M = \max[|A|: A \in \mathscr{A}] \geqq (n+1)/2$, and suppose $A_1, \cdots, A_k$ are the elements of $\mathscr{A}$ of size M. A set of size M has M subsets of size $M - 1$, so a list of all subsets of size $M - 1$ of all the A_i has length kM if repetitions are allowed. But a set can appear in the list at most $n - M + 1$ times, that being the total number of sets of size M that contain it. Hence the number of distinct sets in the list is at least $kM/(n - M + 1)$, which in turn is at least k if $M \geqq (n+1)/2$. Thus it is possible to replace $A_1, \cdots, A_k$ by k distinct subsets of size $M - 1$ without changing $|\mathscr{A}|$ and without destroying the defining property of $\mathscr{A}$.

By repeated application of this procedure, we can ensure that $M = [n/2]$. Similarly, we can ensure that $\min[|A|: A \in \mathscr{A}] = [n/2]$. But if $|A| = [n/2]$ for every A in $\mathscr{A}$, then

$$|\mathscr{A}| \leqq \binom{n}{[n/2]}.$$

Now (11.6) follows by Stirling's formula. □

12. References. The standard treatment of probabilistic number theory is Kubilius' book [22]; introductions can be found in [20] and in [6]. There are extensive bibliographical notes in [22]; for work subsequent to the appearance of [22], see the review paper [16] of Galambos and LeVeque's section on additive functions in *Reviews in Number Theory, 1940–1972*, to be published presently by The American Mathematical Society. Because of these bibliographies, only a few remarks are needed here.

Hardy and Ramanujan [18] and Turán [31] obtained the first laws of large numbers; the results of Section 2 here, in their present generality, are due to Kubilius. Erdös and Kac [13] proved the first central limit theorem for additive functions; the proof in Section 3 is essentially that of [4]. The general theorems on convergence to infinitely divisible laws were proved by Kubilius, who introduced the class H. The moment method of Section 5 is that of Misevičius [25], adapted to arrays $\{f_n\}$, extended to the case of vague convergence (see (5.10)), and simplified by the methods of [4].

The invariance principle of Section 4 is implicit in Theorem 7.3 of [22]; one of the manuscript versions of [3] contained the present proof. The same result under the Lindeberg conditions, Theorem 6.2 above, was proved by Philipp [27] and by Babu [1]. The general theory in Section 6 is new.

The results in Sections 7 and 8 are due to Erdös [10], [12] and to Erdös and Wintner [14]. The proofs are theirs, simplified somewhat.

Section 9 is new.

LeVeque's conjecture that the error term in the Erdös–Kac theorem sohuld be the same as for the classical central limit theorem was proved by Rényi and Turán [28] by methods of analytic number theory. No probabilistic proof has been found; probably the methods of this paper cannot be adapted to that problem.

REFERENCES

[1] BABU, GUTTI JOGESH (1973). Probabilistic methods in the theory of arithmetic functions. Ph. D. dissertation, The Indian Statistical Institute, Calcutta.

[2] BILLINGSLEY, PATRICK (1965). *Ergodic Theory and Information.* Wiley, New York.

[3] BILLINGSLEY, PATRICK (1968). *Convergence of Probability Measures.* Wiley, New York.

[4] BILLINGSLEY, PATRICK (1969). On the central limit theorem for the prime divisor function. *Amer. Math. Monthly* **76** 132–139.

[5] BILLINGSLEY, PATRICK (1971). Weak convergence of measures: Applications in probability. *Regional Conference Series in Applied Mathematics.* SIAM Philadelphia.

[6] BILLINGSLEY, PATRICK (1973). Prime numbers and Brownian motion. *Amer. Math. Monthly* **80** 1099–1115.

[7] CHUNG, KAI LAI (1968). *A Course in Probability Theory.* Harcourt, Brace and World, New York.

[8] CRAMÉR, HARALD (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[9] DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[10] ERDÖS, PAUL (1935, 1937, 1938). On the density of some sequences of numbers. Part I, *J. London Math. Soc.* **10** 120–125; Part II, *ibid.* **12** 7–11; Part III, *ibid.* **13** 119–127.

[11] ERDÖS, PAUL (1939). On the smoothness of the asymptotic distribution of additive arithmetical functions. *Amer. J. Math.* **61** 722–725.

[12] ERDÖS, PAUL (1946). On the distribution of additive functions. *Ann. of Math.* (2) **47** 1–20.

[13] ERDÖS, PAUL and KAC, MARK (1940). The Gaussian law of errors in the theory of additive number-theoretic functions. *Amer. J. Math.* **62** 738–742.

[14] ERDÖS, PAUL and WINTNER, AUREL (1939). Additive arithmetical functions and statistical independence. *Amer. J. Math.* **61** 713–721.

[15] FELLER, WILLIAM (1968, 1971). *An Introduction to Probability Theory and its Applications.* **1** 3rd ed. **2** 2nd ed. Wiley, New York.

[16] GALAMBOS, J. (1970). Distribution of arithmetical functions. A survey. *Ann. Inst. H. Poincaré, Sect. B.* **6** 281–305.

[17] GNEDENKO, B. V. and KOLMOGOROV, A. N. (1954). *Limit Distribution for Sums of Independent Random Variables.* Addison-Wesley, Reading. (Translation of the Russian edition of 1949).

[18] HARDY, G. H. and RAMANUJAN, S. (1920). The normal number of prime factors of a number *n*. *Quarterly J. Math.* **48** 76–92.

[19] HARDY, G. H. and WRIGHT, E. M. (1960). *An Introduction to the Theory of Numbers*, 4th ed. Clarendon Press, Oxford.

[20] KAC, MARK (1959). *Statistical Independence in Probability, Analysis and Number Theory.* Carus Math. Monograph 12. Wiley, New York.

[21] KARLIN, SAMUEL (1966). *A First Course in Stochastic Processes.* Academic Press, New York.

[22] Kubilius, J. (1964). *Probabilistic Methods in the Theory of Numbers*. Amer. Math. Soc. Translations **11**. (Translation of the second Russian edition of 1962.)

[23] Lévy, Paul (1931). Sur les séries dont les termes sont des variables éventuelles indépendantes. *Studia Math*. **3** 119–155.

[24] Loève Michel (1960). *Probability Theory*, 2nd ed. Van Nostrand, Princeton.

[25] Misevičius, G. (1965). The use of the method of moments in probabilistic number theory. *Litovsk. Mat. Sb*. **5** 275–289, (in Russian).

[26] Philipp, Walter (1971). Mixing sequences of random variables and probabilistic number theory. *Mem. Amer. Math. Soc*. **114**.

[27] Philipp, Walter (1973). Arithmetic functions and Brownian motion *Proc. Sympos. Pure Math*. **24** 233–246.

[28] Rényi, A. and Turán, P. (1957). On a theorem of Erdös-Kac. *Acta. Arith*. **4** 71–84.

[29] Scheffé, H. (1947). A useful convergence theorem for probability distributions. *Ann. Math. Statist*. **18** 434–438.

[30] Sperner, Emanuel (1928). Ein Satz über Untermengen einer endlichen Menge. *Math. Zeit*. **27** 544–548.

[31] Turán, Paul (1934). On a theorem of Hardy and Ramanujan. *J. London Math. Soc*. **9** 274–276.

Department of Statistics
The University of Chicago
1118 East 58th Street
Chicago, Illinois 60637