ACTA ARITHMETICA
XXVI (1974)

A probabilistic setting for prime number theory*

by

M. C. WUNDERLICH (DeKalb, Ill.)

Introduction. This paper is generally concerned with questions of stability with respect to theorems in prime number theory. To what extent does a theorem, true for the sequence of primes, remain true when one perturbs the defining characteristics of the primes. The work of Beurling [2] and more recently Bateman and Diamond [1] represent an analytic approach to this question and the proof of the prime number theorem for generalized primes represents a true generalization of the prime number theorem. The author has been interested in questions of stability from a quite different point of view. If one perturbs the defining characteristics of the sieve of Eratosthenes, one obtains a different sequence of "primes" and the author [6] has demonstrated a high degree of stability for the prime number theorem for these sieve generated sequences.

In 1958, David Hawkins [3] described this type of perturbation in terms of the following stochastic process: let A_1 be the sequence $\{2, 3, 4, 5, \ldots\}$. Since $A_1(1) = 2$, we eliminate from A_1 all elements exceeding 2 with probability $\frac{1}{2}$, producing A_2. In general, if $A_n(n) = k$, you eliminate each element in A_n which exceeds k with probability $1/k$, producing A_{n+1}. One would like to determine to what extent the prime number theorem, or any other theorem concerning primes, is true for the sequence $A = \bigcup_{n=1}^{\infty} A_n$ generated by this process.

Recently, David Hawkins [4] constructed a sequence of finite probability spaces P_n which emulated this process. The elements of P_n are sequences of integers $\leqslant n$ and he defines the random variable $h_n(\varrho)$, $\varrho \in P_n$, to be the number of elements in ϱ. He proceeds to obtain statistical estimates for the distributions of h_n and is able to show that the prime number theorem holds in these spaces in the sense of the weak law of large numbers.

* The research for this paper was partially supported by NSF grant GP 23299.

The purpose of this paper is threefold.

A. We construct in Section I a single probability space X which emulates the Hawkins sieve. The finite spaces P_n correspond to cylinder sets in X.

B. In Section II, the result of Hawkins is obtained in X by obtaining similar but sharper statistical estimates for the distributions of h_n.

C. In Section III, we employ a method of Paul Lévy [5] to prove a theorem which enables us to obtain results in the sense of the strong law of large numbers from results in the sense of the weak law of large numbers for sequences of random variables $\{f_n(a)\}$ which possess certain properties. Using this theorem, we are able to generalize Hawkins' result and show that the prime number theorem holds for almost all sequences $a \in X$. We also show that an analog of Merten's theorem holds for almost all sequences $a \in X$; that is to say,

$$\prod_{\substack{a \in a \\ a<n}} \left(1-\frac{1}{a}\right) \sim \frac{1}{\log n}.$$

Finally, we study a sequence of random variables in X which essentially counts the number of consecutive integers in a sequence. We prove that for almost all sequences $a \in X$, the number of consecutive pairs $k-1$, k for which $k < n$ is asymptotic to $n/\log^2 n$. This result is analogous to Hardy and Littlewood's conjecture regarding the density of twin primes.

The author would like to thank Professor David Williams (University of Wales, Swansea) for suggesting the method of approach used in Section III.

I. The probability space. In this first part of the paper, we will construct the probability space. Throughout the section, lower case Greek letters $a, \beta, \gamma, \ldots$ will denote sequences of integers. Upper case Latin letters $A, B, C, \ldots$ denote sets of sequences and upper case script letters $\mathscr{A}, \mathscr{B}, \mathscr{C}, \mathscr{D}, \ldots$ will denote classes of sets of sequences. The integers themselves are denoted by lower case Latin letters. Note that in this section, ε is a sequence and not a small positive real number. If a is a sequence and n an integer greater than 2, we will use a_n to denote the finite sequence $a \cap \{2, 3, 4, \ldots, n-1\}$. (None of our sequences will contain 1.) In other words, a_n is the set of integers of a which are less than n. Correspondingly, a^n denotes the set of integers in a not less than n.

We begin by letting X be the space of all sequences of integers greater than 1. X contains finite as well as infinite sequences. $\mathscr{X}$ is the class of all sets of such sequences, i.e. the power set 2^X. The following definition establishes a class of elementary sets over which we can define a probability measure.

Definition 1. $E \in \mathscr{X}$ is called an *elementary set of sequences*, or just an *elementary set* if there exists a finite sequence $\{a_1, a_2, \ldots, a_k\} \in X$ and an integer $n > a_k$ for which $\varepsilon \in E$ if and only if

$$\varepsilon_n = (a_1, a_2, \ldots, a_k).$$

E is denoted $(a_1, a_2, \ldots, a_k; n)$ or $(a; n)$ if a is being used to denote the sequence $\{a_1, a_2, \ldots, a_k\}$. N is referred to as the *order* of the elementary set. Note that k could be zero in which case $(\ ; n)$ is the set of all sequences whose elements are not less than n. We will assume that $n \geq 2$ since 1 is never an element of our sequences. Note also that $(\ ; 2) = X$. The class of elementary sets is denoted by $\mathscr{E}$.

Our probability function is now defined recursively on $\mathscr{E}$.

Definition 2. μ is a real valued function defined on $\mathscr{E}$ to satisfy

(a) $\mu(\ ; 2) = 1$.

(b) $\mu(a_1, a_2, \ldots, a_k, n; n+1) = \mu(a_1, a_2, \ldots, a_k; n) \prod_{i=1}^{k} \left(1-\frac{1}{a_i}\right)$.

(c) $\mu(a_1, a_2, \ldots, a_k; n+1) = \mu(a_1, a_2, \ldots, a_k; n)\left(1-\prod_{i=1}^{k}\left(1-\frac{1}{a_i}\right)\right)$.

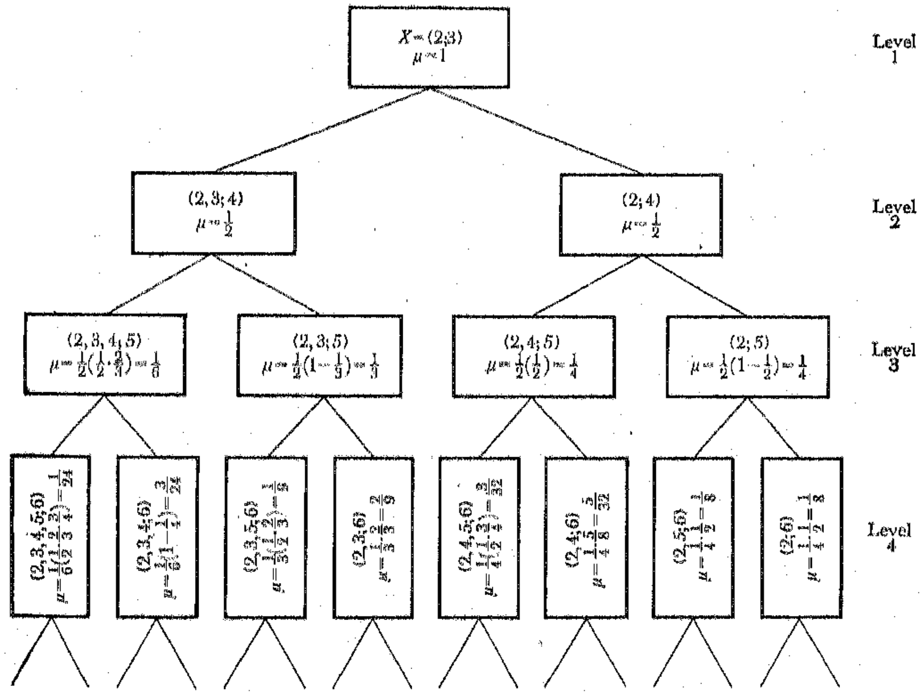

Fig. 1

Note that (c) with $k=0$ and $n=2$ implies that the class of sequences not containing 2 has probability zero (assuming that empty products equal 1).

This definition is perhaps better understood by regarding it as a binary tree. (See Figure 1.) The root of the tree is $X=(2;\ 3)$ which has probability 1. It is divided into two subclasses — those sequences containing 3 and those that do not. Each subclass has, in this case, probability $\frac{1}{2}$. Each of these classes is separated into two classes as shown in the figure. Thus each node B of the tree corresponds to an elementary set $(a_1, a_2, \ldots, a_k; n)$ and is divided into two elementary sets B_1 and B_2 depending on whether or not n is or is not in the sequences.

The probabilities of these subclasses are chosen to emulate the Hawkins sieve: n will have been eliminated with the probabilities $1/2, 1/a_2, 1/a_3, \ldots, 1/a_k$, so that the conditional probability of its remaining is $\prod_{i=1}^{k}(1-1/a_k)$. Hence we obtain $\mu(B_1)$ and $\mu(B_2)$ by multiplying $\mu(B)$ by $\prod_{i=1}^{k}(1-1/a_i)$ and $1-\prod_{i=1}^{k}(1-1/a_i)$ respectively.

This tree structure will be useful later in the proof of several lemmas. Note that when A is below B on the tree, $A\subset B$. If A is neither above or below B, $A\cap B=\emptyset$. The level of each node corresponds exactly with $n-2$.

We will now want to extend the definition of μ to a ring of subsets of X and to do this we need the following property:

LEMMA 1. *μ is finitely additive on $\mathscr{E}$, that is, if $A=B_1\cup B_2\cup\ldots\cup B_k$ where A and the B_i are in $\mathscr{E}$ and the B_i are disjoint, then*

$$\mu(A)=\sum_{j=1}^{k}\mu(B_j).$$

Proof. We will utilize the binary tree structure of μ to formulate an induction argument on k. If $k=2$, then A must be of the form diagrammed to the right and the lemma follows from Definition 2. Assume now that the lemma holds for $k<m$ and assume that $A=B_1\cup\cup B_2\cup\ldots\cup B_m$. Since the B's are a finite set, two of them, say B_{m-1} and B_{m-2} must be subtrees of the same node, say B'. (Otherwise the union would not be an elementary set.) By definition, we can replace $B_m\cup B_{m-1}$ by B' and our inductive assumption completes the argument.

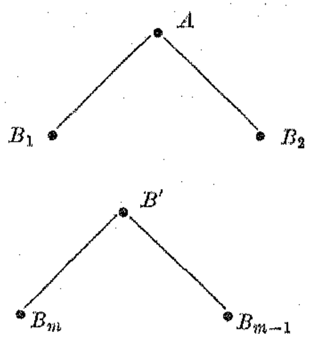

Fig. 2

DEFINITION 3. A set of sequences R is *eventually arbitrary* if a positive integer $n>1$ exists which has the following property: For any sequence $\varrho\in R$ and for any sequence $\eta_i=(n_1, n_2, \ldots)$ of integers $n_i\geqslant n$, there exists a sequence $\bar{\varrho}\in R$ such that $\bar{\varrho}_n=\varrho_n$ and $\bar{\varrho}^n=\eta$. In terms of elementary sets, we say that R is eventually arbitrary if for every $\varrho\in R$ where $\varrho_n=\{r_1, r_2, \ldots, r_k\}$, the elementary set $(r_1, r_2, \ldots, r_k; n)$ is contained in R. Clearly if this condition holds for n, it holds for any $m>n$, so we define the minimal such n to be the *order* of the eventually arbitrary set. We denote by $\mathscr{R}$ the class of all eventually arbitrary sets.

LEMMA 2. *$\mathscr{R}$ is an algebra of subsets.*

Proof. A set $R\subset\mathscr{R}$ can be represented as a finite disjoint union of elementary sets in the following way: if n is the order of R, let $a_1, a_2, a_3, \ldots, a_k$ be the set of all distinct finite sequences ϱ_n where ϱ ranges over all of R. Then $\bigcup_{i=1}^{k}(a_i;\ n)=R$ by Definition 3. Also, any finite union of elementary sets is eventually arbitrary. Thus R is closed under finite unions. To show that $\mathscr{R}$ is closed under differences we must show that if E_i and E_j are in $\mathscr{E}$, then E_i-E_j is in $\mathscr{R}$. One can see from the binary tree representation that either $E_i\subset E_j$, $E_j\subset E_i$, or $E_i\cap E_j=\emptyset$. Thus we need only consider the case $E_j\subset E_i$. Here we can express E_i as a union of all elementary sets contained in E_i at the level (on the tree) of E_j and thus E_i-E_j would be that same union minus E_j. Finally, X is clearly eventually arbitrary making $\mathscr{R}$ an algebra.

We can now extend μ to $\mathscr{R}$ by representing each element R as a disjoint union $E_1, E_2, \ldots, E_k$ and defining $\mu(R)=\sum_{i=1}^{k}\mu(E_i)$. Lemma 1 can be used to show that the definition is well defined. We have then, at the moment, a finitely additive set function μ defined on an algebra of sets $\mathscr{R}$ such that $\mu(X)=1$. In order to use the measure extension theorem we must prove that μ is countably additive on $\mathscr{R}$, that is

$$\mu\Big(\bigcup_{i=1}^{\infty}R_i\Big)=\sum_{i=1}^{\infty}\mu(R_i) \tag{1}$$

whenever the union is disjoint and contained in $\mathscr{R}$. We will show that (1) holds vacuously, in that countable disjoint unions of elements in $\mathscr{R}$ cannot themselves be in $\mathscr{R}$. Since an element of $\mathscr{R}$ can be expressed as a finite disjoint union of elementary sets, $R=\bigcup_{i=1}^{\infty}R_i$ would imply

$$\bigcup_{i=1}^{k}E_i=\bigcup_{j=1}^{\infty}\bar{E}_j$$

and the pigeonhole principle would imply that at least one E_i contains as subsets a countable union of disjoint elementary sets.

LEMMA 3. *If E is an elementary set and $E_1, E_2, \ldots$ is an infinite mutually disjoint set of elementary sets contained in E, then there exists a sequence $a \in E$ not contained in $\bigcup_j E_j$.*

Proof. In Figure 3, the binary tree corresponding to the elementary set E is depicted and the vertices corresponding to the sets $E_1, E_2, \ldots$ are circled. To prove the theorem, delete any vertex if it is an ancestor

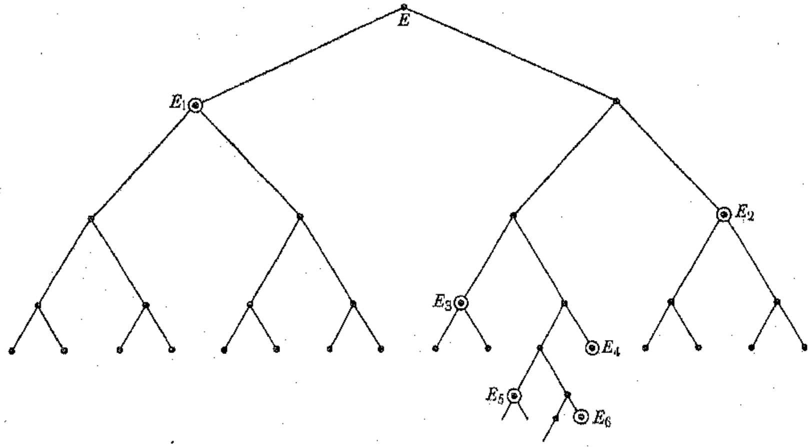

Fig. 3

of one of the E_i. If the remaining tree does not contain an infinite path from the root downwards, it is finite and the sequence $E_1, E_2, E_3, \ldots$ consists of the terminal nodes which are finite in number contrary to the hypothesis of the lemma. Such an infinite path clearly represents a sequence of integers not in any of the E_i.

We have now shown that μ is a countably additive set function on an algebra, and we can use the well known extension theorem to extend it to the smallest σ-algebra containing $\mathcal{R}$. We will denote this class by $\mathcal{S}$.

There are two important classes of sets, contained in $\mathcal{S}$ which are of interest to us for this paper. Let f be a positive non-decreasing real-valued function and let S_f be the set of all sequences $\{a_1, a_2, \ldots\}$ satisfying $\sum_{a_k \leqslant x} 1 \leqslant f(x)$. Similarly, we let S^f be the set of sequences that satisfy $\sum_{a_k \leqslant x} 1 \geqslant f(x)$. These are elements of $\mathcal{S}$ since they can be represented as an intersection of a countable number of eventually arbitrary sets.

II. In this section, we will show that the sequences contained in the probability space tend to be prime-like in that the number of elements less than x approximates $x/\log x$ in the sense of the weak law of large numbers. Except for Lemma 4, all the results in this section were obtained

by David Hawkins [4] in a slightly different but logically equivalent context. Also, (31) gives a sharper estimate for the variance than that obtained by Hawkins.

We begin by defining two sequences of random variable on X.

DEFINITION 4. If $n \geqslant 2$ and $a \in X$ then define $x_n(a)$ and $h_n(a)$ by

$$x_n(a) = \prod_{\substack{a \in a \\ a < n}} \left(1 - \frac{1}{a}\right), \quad x_2(a) = 1,$$

$$h_n(a) = \sum_{\substack{a \in a \\ a < n}} 1.$$

Since $x_n(a)$ and $h_n(a)$ have constant values on any elementary set A_n of order n, we will often employ the notation $x(A_n)$ and $h(A_n)$ to represent $x_n(a)$ and $h_n(a)$ for $a \in A_n$. If we let

$$z_n(a) = \begin{cases} x_n(a) & \text{if} \quad n \in a, \\ 1 - x_n(a) & \text{if} \quad n \notin a, \end{cases} \tag{2}$$

then the following lemma gives us an explicit expression for the measure of any elementary set.

LEMMA 4. *If a is any sequence in an elementary set A_n of order n, then*

$$\mu(A_n) = \prod_{j=3}^{n-1} z_j(a).$$

The proof is easily obtained by iterating in the definition of μ.

LEMMA 5. *If we let S_n represent the set of all sequences containing n, then*

$$\mu(S_n) = E[x_n] = \int_X x_n \, d\mu. \tag{3}$$

Proof. It follows from Definition 2 that we can write

$$\mu(S_n) = \sum_{B_n} \mu(B_n) x(B_n)$$

where the sum is extended over all elementary sets B_n of order n (a notation which we will use frequently) but this is exactly the expectation required in (3).

We now use Definition 2 to obtain the following recurrence relation for the kth moment of x_n.

$$\begin{aligned} E[x_{n+1}^k] &= \sum_{B_n} \mu(B_n)\big(1 - x(B_n)\big) x^k(B_n) + \mu(B_n) x(B_n) x^k(B_n) \left(1 - \frac{1}{n}\right)^k \\ &= \sum_{B_n} \mu(B_n) \left(\left(1 - \frac{1}{n}\right)^k - 1\right) x^{k+1}(B_n) + \mu(B_n) x^k(B_n) \end{aligned} \tag{4}$$

which is equivalent to

(5) $$E[x_{n+1}^k]-E[x_n^k] = \left(\left(1-\frac{1}{n}\right)^k - 1\right)E[x_n^{k+1}].$$

We observe that the above relation holds for all integral k and is subject to the boundary condition $E[x_2^k] = 1$. Letting $k = -1$ in (5) and summing from 2 to $n-1$, we obtain

(6) $$E[x_n^{-1}] = 1+\sum_{k=1}^{n-2}\frac{1}{k} = \log(n-2)+\gamma+1+O\left(\frac{1}{n}\right).$$

To obtain a relationship between $E[x_n]$ and $E[x_n^{-1}]$, we let $k = 1$ in (5) and write it in the form

(7) $$\frac{1}{E[x_{n+1}]}-\frac{1}{E[x_n]} = \left[\frac{nE^2[x_n]}{E[x_n^2]}-E[x_n]\right]^{-1}.$$

If we apply Schwartz's inequality to the random variable $x_n^{\frac{k-1}{2}}$ and $x_n^{\frac{k+1}{2}}$, we obtain

(8) $$E[x_n^{k+1}]E[x_n^{k-1}] \geqslant E^2[x_n^k]$$

and thus

$$E[x_n^2] \geqslant E^2[x_n].$$

In (7) this implies

$$E^{-1}[x_{n+1}]-E^{-1}[x_n] \geqslant \frac{1}{n-E[x_n]} \geqslant \frac{1}{n}$$

which when summed from 2 to $n-1$ gives

(9) $$E^{-1}[x_n] \geqslant \log(n-1)+\gamma+O\left(\frac{1}{n}\right).$$

But Schwartz's inequality also implies that $E[x_n]E[x_n^{-1}] \geqslant 1$ which together with (9) and (6) has proved

LEMMA 6.

$$\mu(S_n) = \frac{1}{\log(n-1)}+O\left(\frac{1}{\log^2(n-1)}\right).$$

The following lemma is a straight forward generalization of the previous result.

LEMMA 7.

$$E[x_n^k] = \frac{1}{\log^k(n-1)}+O\left(\frac{1}{\log^{k+1}(n-1)}\right) \quad \textit{for any positive } k.$$

Proof. We use induction on k and use as our inductive assumption the statement of the lemma together with the statement

(10) $$E[x_n^{-k}] = \log^k(n-1)+O\left(\log^{k-1}(n-1)\right).$$

Lemma 6 and (6) verify the truth of the inductive assumption for $k = 1$, so we assume it true for $k-1$ and proceed by using (5) to obtain

$$\begin{aligned} E[x_{n+1}^{-k}]-E[x_n^{-k}] &= \left(\left(1+\frac{1}{n-1}\right)^k-1\right)E[x_n^{-k+1}] \\ &= \left(\frac{k}{n-1}+O\left(\frac{1}{n^2}\right)\right)E[x_n^{-k+1}] \\ &= \frac{k\log^{k-1}(n-1)}{n-1}+O\left(\frac{\log^{k-2}(n-1)}{n-1}\right). \end{aligned}$$

Summing this from 2 to $n-1$ verifies (10). We now use (5) and write

$$\begin{aligned} \frac{1}{E[x_{n+1}^k]}-\frac{1}{E[x_n^k]} &= \left[\frac{E[x_{n+1}^k]E[x_n^k]}{E[x_n^k]-E[x_{n+1}^k]}\right]^{-1} \\ &= \left[\frac{E^2[x_n^k]-\left(\frac{k}{n}+O\left(\frac{1}{n^2}\right)\right)E[x_n^{k+1}]E[x_n^k]}{\left(\frac{k}{n}+O\left(\frac{1}{n^2}\right)\right)E[x_n^{k+1}]}\right]^{-1} \\ &\geqslant \left[\frac{E^2[x_n^k]}{\left(\frac{k}{n}+O\left(\frac{1}{n^2}\right)\right)E[x_n^{k+1}]}\right]^{-1} \geqslant \frac{\frac{k}{n}+O\left(\frac{1}{n^2}\right)}{E[x_n^{k-1}]} \\ &= \frac{k}{n}\log^{k-1}(n-1)+O\left(\log^{k-2}(n-1)\right). \end{aligned}$$

We have made use of the inductive assumption and (8). Summing this from 2 to $n-1$ yields

$$E^{-1}[x_n^k] \geqslant \log^k n+O(\log^{k-1}n)$$

which together with (10) and the fact that $E[x_n^k]E[x_n^{-1}] \geqslant 1$ (Schwartz) completes the proof of the lemma.

We now consider the identity

$$Mx = \frac{1}{1+\left(\frac{x^{-1}-M}{M}\right)}$$

$$= 1 - \frac{(x^{-1}-M)}{M} + \frac{(x^{-1}-M)^2}{M^2} - \ldots + (-1)^l\left\{\frac{x(x^{-1}-M)^l}{M^{l-1}}\right\}$$

or

(11) $$x = \frac{1}{M} - \frac{(x^{-1}-M)}{M^2} + \frac{(x^{-1}-M)^2}{M^3} - \ldots + (-1)^l\left\{\frac{x(x^{-1}-M)^l}{M^l}\right\}.$$

If we let $M = M_n = E[x_n^{-1}]$, let $x = x_n$ and integrate (11) over X we obtain

(12) $$E[x_n] = \frac{1}{M_n} + \frac{E[(x_n^{-1}-M_n)^2]}{M_n^3} - \ldots + (-1)^l\frac{E[x(x_n^{-1}-M_n)^l]}{M_n^l}.$$

Raising (11) to the kth power before integrating produces

(13) $$E[x_n^k] = \frac{1}{M_n^k} + \binom{k+1}{2}\frac{E[(x_n^{-1}-M_n)^2]}{M_n^{k+2}} - \binom{k+2}{3}\frac{E[(x_n^{-1}-M_n)^3]}{M_n^{k+3}} + \ldots + (-1)^{kl}\frac{E[x^k(x_n^{-1}-M_n)^{kl}]}{M_n^{kl}}.$$

Lemma 8. *There exists positive constants $c_{i,j}$ such that for each i, a c_i exists such that*

(14) $$\lim_{j\to\infty} c_{i,j} = c_i$$

and

(15) $$E[x_n] = \frac{1}{M_n} + \frac{c_{2,n}}{M_n^3} - \frac{c_{3,n}}{M_n^4} + \ldots + O\left(\frac{1}{M_n^{l+1}}\right)$$

and

(16) $$E[x_n^k] = \frac{1}{M_n^k} + \binom{k+1}{2}\frac{c_{2,n}}{M_n^{k+2}} - \binom{k+2}{3}\frac{c_{3,n}}{M_n^{k+3}} + \ldots + O\left(\frac{1}{M_n^{kl+1}}\right).$$

Proof. In view of (12) and (13) we need only verify (14), that is, the central moments

$$c_{k,n} = E[(x_n^{-1}-M_n)^k]$$

tend to finite limits as $n\to\infty$.

We observe first that

(17) $$c_{k,n} = \sum_{s=0}^{k}(-1)^s\binom{k}{s}M_n^{k-s}E[x_n^{-s}].$$

If we let $\Delta c_{k,n} = c_{k,n+1}-c_{k,n}$ (13) will follow if we can prove that $\Delta c_{k,n} = O(1/n^{1+\varepsilon})$ for a fixed positive ε. For notational convenience, we will write $a \doteq b$ to mean $a = b+O(1/n^{1+\varepsilon})$ thus permitting us to add and subtract terms at will which are $O(1/n^{1+\varepsilon})$. Since $M_{n+1}^q = M_n^q + \frac{q}{n-1}M_n^{q-1} + O(1/n^{1+\varepsilon})$ and since $E[x_n^{-q}] = O(\log^q n)$ we can write, using (17),

$$\Delta c_{k,n} = \sum_{s=0}^{k}(-1)^s\binom{k}{s}\left(M_{n+1}^{k-s}E[x_{n+1}^{-s}] - M_n^{k-s}E[x_n^{-s}]\right)$$

$$\doteq \sum_{s=0}^{k}(-1)^s\binom{k}{s}M_{n+1}^{k-s}\left(E[x_{n+1}^{-s}]-E[x_n^{-s}]\right) + \sum_{s=0}^{k-1}(-1)^s\binom{k}{s}\frac{k-s}{n-1}M_n^{k-s-1}E[x_n^{-s}]$$

$$= A+B.$$

Now using (5),

$$A = \sum_{s=0}^{k}(-1)^s\binom{k}{s}M_{n+1}^{k-s}\left(\left(1+\frac{1}{n-1}\right)^s-1\right)E[x_n^{-s+1}]$$

$$\doteq \sum_{s=0}^{k}(-1)^s\binom{k}{s}M_{n+1}^{k-s}\left(\frac{s}{n-1}\right)E[x_n^{-s+1}]$$

$$= \sum_{s=0}^{k-1}(-1)^{s+1}\binom{k}{s+1}\left(\frac{s+1}{n-1}\right)M_{n-1}^{k-s-1}E[x_n^{-s}]$$

$$= \sum(-1)^{s+1}\binom{k}{s}\left(\frac{k-s}{n-1}\right)M_{n-1}^{k-s-1}E[x_n^{-s}]$$

$$= -B.$$

To complete the proof of the lemma, we must verify that the last terms of (12) and (13) are finite. The Schwartz inequality yields

$$E^2[x_n(x_n^{-1}-M_n)^l] \leqslant E[x_n^2]E[(x_n^{-1}M_n)^{2l}]$$

which converges by Lemma 7. Similarly for the last term of (13), we write

$$E^2[x^k(x_n^{-1}-M_n)^{kl}] \leqslant E[x^{2k}]E[(x_n^{-1}-M_n)^{2kl}]$$

and again use Lemma 7.

We now turn our attention to the random variable $h_n(a)$. Our objective will be to compute its expected value and its variance and thus be in a position to use Chebyshev's inequality. The fundamental recurrence relation for h^k is

$$E[h_{n+1}^k] = \sum_{B_n} \mu(B_n)\big(1-x(B_n)\big)\, h^k(B_n) + \mu(B_n)x(B_n)\big(h(B_n)+1\big)^k$$

from which we obtain

(18) $$E[h_{n+1}] - E[h_n] = E[x_n]$$

and

(19) $$E[h_{n+1}^2] - E[h_n^2] = 2E[x_n h_n] + E[x_n].$$

We can similarly obtain

$$E[h_{n+1}x_n^k] = \sum_{B_n} \mu(B_n)\big(1-x(B_n)\big)h(B_n)x^k(B_n) + $$
$$+\mu(B_n)x(B_n)\big(h(B_n)+1\big)x^k(B_n)\left(1-\frac{1}{n}\right)^k$$

from which we obtain

$$E[h_{n+1}x_{n+1}^k] - E[h_n x_n^k]$$
$$= \left(1-\frac{1}{n}\right)^k E[x_n^{k+1}] - \frac{k}{n}E[h_n x_n^{k+1}] + O\left(\frac{1}{n^2}E[h_n x_n^{k+1}]\right).$$

However since $h_n < n$, we can sum the above expression obtaining $E[h_n x_n^k] = O(nE[x_n^{k+1}])$ and thus obtain the simplified expression

(20) $$E[h_{n+1}x_{n+1}^k] - E[h_n x_n^k] = E[x_n^{k+1}] - \frac{k}{n}E[h_n x_n^{k+1}] + O\left(\frac{1}{n}E[x_n^{k+1}]\right).$$

LEMMA 9. *If r, s and t are integers satisfying $s \geqslant 0$ and $0 < r < t$, then*

(21) $$\sum_{k=2}^{n} \frac{k^s}{M_k^r} = \frac{1}{s+1}\cdot\frac{n^{s+1}}{M_n^r} + \frac{c(1,r,s)n^{s+1}}{M_n^{r+1}} + \ldots + $$
$$+\frac{c(t-r-1,r,s)n^{s+1}}{M_n^{r+(t-r-1)}} + O\left(\frac{n^{s+1}}{M_n^t}\right)$$

where

$$c(1,r,s) = \frac{r}{(s+1)^2}$$

and

(22) $$c(i,r,s) = \frac{r(r+1)\ldots(r+i-1)}{(s+1)^{i+1}}.$$

Proof. If we let $r = t-1$, (21) becomes

$$\sum_{k=2}^{n} \frac{k^s}{M_k^{t-1}} = \frac{1}{s+1}\frac{n^{s+1}}{M_n^{t-1}} + O\left(\frac{n^{s+1}}{M_n^t}\right)$$

which follows from (6). Assuming that (21) holds for $r+1$, $r < t-1$, partial summation yields

$$\sum_{k=2}^{n} \frac{k^s}{M_k^r} = \frac{1}{s+1}\frac{n^{s+1}}{M_n^r} + \sum_{k=2}^{n-1} \frac{k^{s+1}}{s+1}\left(\frac{M_{k+1}^r - M_k^r}{M_k^r M_{k+1}^r}\right).$$

Since $M_{k+1} - M_k = \dfrac{1}{k-1}$, we have for any $\varepsilon > 0$

(22.5) $$M_{k+1}^r = M_k^r + r\frac{M_k^{r-1}}{k-1} + O\left(\frac{1}{k^{2-\varepsilon}}\right)$$

so that

$$\sum_{k=2}^{n} \frac{k^s}{M_k^r} = \frac{1}{s+1}\left(\frac{n^{s+1}}{M_n^r} + r\sum_{k=2}^{n-1}\frac{k^s}{M_k^{r+1}} + O(n^{s+\varepsilon})\right)$$
$$= \frac{1}{s+1}\left(\frac{n^{s+1}}{M_n^r} + \frac{r}{s+1}\frac{n^{s+1}}{M_k^{r+1}} + \frac{rc(1,r+1,s)}{s+1}\frac{n^{s+1}}{M_k^{r+2}} + \right.$$
$$\left. + \ldots + \frac{rc(t-r,r+1,s)}{s+1}\frac{n^{s+1}}{M_k^{r+1+(t-r)}} + O\left(\frac{n^{s+1}}{M_k^t}\right)\right).$$

However, (22) implies that $c(i,r,s) = \dfrac{c(i-1,r+1,s)\cdot r}{s+1}$ and this substitution obtains (21). Thus we have obtained (22) for all $0 < r < t$ by reverse induction.

This lemma can be used to obtain the following estimates:

(23) $$\sum_{k=2}^{n} \frac{1}{M_k^4} = \frac{n}{M_n^4} + O\left(\frac{m}{M_k^5}\right),$$

(24) $$\sum_{k=2}^{n} \frac{1}{M_k^3} = \frac{n}{M_n^3} + \frac{3n}{M_n^4} + O\left(\frac{n}{M_n^5}\right),$$

(25) $$\sum_{k=2}^{n} \frac{1}{M_k^2} = \frac{n}{M_n^2} + \frac{2n}{M_n^3} + \frac{6n}{M_n^4} + O\left(\frac{n}{M_n^5}\right),$$

(26) $$\sum_{k=2}^{n} \frac{1}{M_k} = \frac{n}{M_n} + \frac{n}{M_n^2} + \frac{2n}{M_n^3} + \frac{6n}{M_n^4} + O\left(\frac{n}{M_n^5}\right),$$

(27) $$\sum_{k=2}^{n}\frac{k}{M_k^4} = \frac{1}{2}\frac{n^2}{M_n^4}+O\left(\frac{n^2}{M_n^5}\right),$$

(28) $$\sum_{k=2}^{n}\frac{k}{M_k^3} = \frac{1}{2}\frac{n^2}{M_n^3}+\frac{3}{4}\frac{n^2}{M_n^4}+O\left(\frac{n^2}{M_n^5}\right),$$

(29) $$\sum_{k=2}^{n}\frac{k}{M_k^2} = \frac{1}{2}\frac{n^2}{M_n^2}+\frac{1}{2}\frac{n^2}{M_n^3}+\frac{3}{4}\frac{n^2}{M_n^4}+O\left(\frac{n^2}{M_n^5}\right),$$

(30) $$\sum_{k=2}^{n}\frac{k}{M_k} = \frac{1}{2}\frac{n^2}{M_n}+\frac{1}{4}\frac{n^2}{M_n^2}+\frac{1}{4}\frac{n^2}{M_n^3}+\frac{3}{8}\frac{n^2}{M_n^4}+O\left(\frac{n^2}{M_n^5}\right).$$

We now sum (from 2 to $n-1$) the expression (20) letting k be 4, 3, 2 and finally 1 using (15) and the estimates (23) through (30) to obtain in turn

$$E[h_n x_n^4] = O(nE[x_n^5]),$$

$$E[h_n x_n^3] = \frac{n}{M_n^4}+O\left(\frac{n}{M_n^5}\right),$$

$$E[h_n x_n^2] = \frac{n}{M_n^3}+\frac{n}{M_n^4}+O\left(\frac{n}{M_n^5}\right),$$

$$E[h_n x_n] = \frac{n}{M_n^2}+\frac{n}{M_n^3}+\frac{n(3C+2)}{M_n^4}+O\left(\frac{n}{M_n^5}\right),$$

where $C = c_{2,n} = \lim\limits_{n\to\infty} E[(x_n^{-1}-M_n)^2]$ in (15). Summing (19) in the same way obtains

$$E[h_n^2] = \frac{n^2}{M_n^2}+\frac{2n^2}{M_n^3}+\frac{n^2(3C+5)}{M_n^4}+O\left(\frac{n^2}{M_n^5}\right)$$

and summing (18) yields

$$E[h_n] = \frac{n}{M_n}+\frac{n}{M_n^2}+\frac{n(C+2)}{M_n^3}+O\left(\frac{n}{M_n^4}\right).$$

Thus the variance $V[h_n]$ can be computed to be

(31) $$V[h_n] = E[h_n^2]-E^2[h_n] = \frac{Cn^2}{M_n^4}+O\left(\frac{n^2}{M_n^5}\right).$$

To estimate C, let

$$C_n = E[(x_n^{-1}-M_n)^2] = E[x_n^{-2}]-M_n^2.$$

Using (5), we can write

$$C_{n+1}-C_n = \frac{1}{(n-1)^2}(M_n-1)$$

and summing this from 2 to $n-1$ yields

$$C_n = \sum_{k=2}^{n-1}\frac{1}{(k-1)^2}(M_k-1) = \sum_{k=3}^{n-1}\frac{1}{(k-1)^2}\sum_{j=1}^{k-2}\frac{1}{j}.$$

S. L. Segal has shown that

$$C = \lim_{n\to\infty} C_n = \zeta(3) = 1.2020569032\ldots$$

We have thus shown that

(32) $$\lim_{n\to\infty} V[h_n] = \frac{Cn^2}{(\log n)^4}.$$

A similar computation can be carried out to show that the third moment of h satisfies

$$E[(h_n-E_n)^3] = \frac{1}{2}\frac{n^3}{M_n^4}+O\left(\frac{n^3}{M_n^5}\right)$$

and hence, in view of (32), all the higher moments of h_n tend to infinity with n.

THEOREM 1 (D. Hawkins). *If $\psi(n)$ is any function of n satisfying*

(33) (a) $\psi(n) = o(\log n)$,
(b) $\psi(n)\to\infty$ *as* $n\to\infty$,

then

(34) $$\mu\left\{a\in X:\ |h_n(a)-E[h_n]| < \frac{n\psi(n)}{\log^2 n}\right\} > 1-O\left(\frac{1}{\psi^2(n)}\right) = 1+o(n).$$

Proof. We apply Chebyshev's inequality using (32).

This shows that the prime number theorem holds in X in the sense of the weak law of large numbers. We can actually say a bit more. If we use the estimate

$$E[h_n] = \frac{n}{\log n}+\frac{n}{\log^2 n}+O\left(\frac{n}{\log^3 n}\right)$$

we can write (34) in the form

(35) $$\mu\left\{a:\ \left|\frac{h_n(a)-\dfrac{n}{\log n}}{\dfrac{n}{\log^{2-\varepsilon} n}}\right| < \frac{\psi(n)}{\log^{\varepsilon} n}\right\} > 1-\frac{C}{\psi^2(n)}.$$

This shows that the error term is smaller than $n/\log^{2-\varepsilon}n$ for any $\varepsilon>0$ in the sense of the weak law of large numbers. If we let $\varepsilon=0$ in (35) we get no useful information.

It would seem reasonable to conjecture that given $\varepsilon>0$, a function $\lambda(n)=\lambda_\varepsilon(n)$ exists for which $\lambda(n)\to\infty$ and such that the set of all sequences $a\in X$ satisfying

$$\left|\frac{h_n(x)-\dfrac{n}{\log n}}{\dfrac{n}{\log^{2+\varepsilon}n}}\right|>\lambda(n)$$

has a measure which tends to one as $n\to\infty$. However, not enough is known about the distribution of h to obtain such a result.

III. In this section, we will employ a method of Paul Lévy [5] to obtain a number of results in the sense of the strong law of large numbers. Our immediate objective is to generalize Theorem 1 and prove that almost all sequences in X satisfy the prime number theorem and a direct proof of this can be found in [7]. However in this paper we will present the method in a more general form so we can apply it to a number of random variables.

THEOREM 2. *Suppose $f_n(a)$, $n=1,2,\ldots$, is a sequence of random variables defined on X having mean and variance $E[f_n]$ and $V[f_n]$ respectively. Suppose furthermore that we can choose functions*

$$E(n)\sim E[f_n],\qquad V(n)=O(V[f_n])\quad \text{and}\quad R(n)=V(n)/E^2(n)$$

which satisfy the conditions

(a) $\lim\limits_{a\to1}\left(\lim\limits_{n\to\infty}\dfrac{E(a^n)}{E(a^{n+1})}\right)=1$ *where* $E(x)=E([x])$,

(b) $R(n)$ *is monotonically non-increasing and*

$$\sum_{k=1}^{\infty}\frac{R(k)}{k}<\infty,$$

(c) $f_n(a)$ *is monotonic in n for any fixed sequence $a\in X$,*

then for almost all sequences $a\in X$, $f_n(a)\sim E(n)$.

It is apparent that the two random variables $x_n(a)$ and $h_n(a)$ which were studied extensively in the previous section satisfy the hypothesis of Theorem 2. In fact, one obtains from (16) that

$$V[x_n]=\frac{1}{\log^4 n}+O\left(\frac{1}{\log^5 n}\right)$$

so that we can choose for both random variables the function

$$R(n)=\frac{1}{\log^2 n}$$

which certainly satisfies condition (b). $h_n(a)$ is monotonically non-decreasing since h is a counting function and $x_n(a)$ is monotonically non-increasing since x is a product of terms less than 1. Thus after proving Theorem 2, we will also have proved

THEOREM 3. *Almost all sequences $a\in X$ satisfy*

$$h_n(a)\sim n/\log n, \tag{36}$$

$$x_n(a)=\prod_{\substack{k<n\\ k\in a}}\left(1-\frac{1}{k}\right)\sim\frac{1}{\log n}. \tag{37}$$

(36) is our probabilistic analog of the prime number theorem and (37) is a probabilistic analog of Mertens' theorem. We should point out at this point that for the sequence of primes, Mertens' theorem produces the constant $e^{-\gamma}$ rather than 1.

Proof of Theorem 2. We will choose any $\delta>0$ and define $l_\delta(n)=l(n)=[(1+\delta)^n]$. If we let $\psi(n)=o(n)$, we can use Chebyshev's inequality to write

$$\mu\{a\in X:\ |f_{l(n)}(a)-E_{l(n)}|>\psi(l(n))E_{l(n)}\}\leqslant\frac{R_{l(n)}}{\psi^2(l(n))}. \tag{38}$$

If we assume that for $x>1$, $R_x=R_{[x]}$, we can use (b) to write

$$\sum_{n=1}^{\infty}R_{l(n)}=\sum_{n=1}^{\infty}R_{a^n}=\sum_{n=1}^{\infty}a^n\left(\frac{R_{a^n}}{a^n}\right)<\sum_{n=1}^{\infty}\frac{2a}{a-1}\sum_{a^{n-1}\leqslant k<a^n}\frac{R_{a^n}}{a^n}$$
$$<\frac{2a}{a-1}\sum_{k=1}^{\infty}\frac{R_k}{k}<\infty;\qquad a=1+\delta.$$

Thus, we can choose a function $\psi(n)=o(n)$ for which the sum of the right hand side of (38) converges. Thus, using the Borel Cantelli lemma, we can assert that for almost all sequences $a\in X$,

$$f_{l(n)}(a)\sim E_{l(n)}. \tag{39}$$

We will complete the proof for the case where $f_n(a)$, and thus $E[f_n]$, is monotonically non-decreasing in n. We will therefore assume that we have chosen E_n to be monotonically non-decreasing in n. If the integer i satisfies $l(n)<i\leqslant l(n+1)$ we can write

$$f_{l(n)}(a)<f_i(a)\leqslant f_{l(n+1)}(a)$$

or

$$\frac{f_{l(n)}(\alpha)}{E_{l(n+1)}} \leqslant \frac{f_i(\alpha)}{E_i} \leqslant \frac{f_{l(n+1)}(\alpha)}{E_{l(n)}}.$$

But then from (39), we have for almost all $\alpha \in X$,

$$(40) \qquad \frac{(1+o(1))E_{l(n)}}{E_{l(n+1)}} \leqslant \frac{f_i(\alpha)}{E_i} \leqslant \frac{(1+o(1))E_{l(n+1)}}{E_{l(n)}}.$$

We now choose a sequence δ_i such that $\delta_i \to 1^+$ and that for $l_i(n) = [(1+\delta_i)^n]$

$$\lim_{n\to\infty} \frac{E_{l_i(n)}}{E_{l_i(n+1)}} = \varrho_i$$

exists. Such a choice is possible by (a). Then letting $\bar{\varrho}_i = \varrho_i + 1/i$, we have from (40) that for almost all sequences $\alpha \in X$ an integer $M = M(\alpha)$ exists such that for $i > M$

$$(41) \qquad \bar{\varrho}_i < \frac{f_i(\alpha)}{E_i} \leqslant \frac{1}{\bar{\varrho}_i}.$$

We will let T_i be the set of all sequences which satisfy (41). Since $\bar{\varrho}_i \to 1$ as $i \to \infty$, the intersection

$$T = \bigcap_{i=1}^{\infty} T_i$$

is exactly the set of all sequences for which $f_i(\alpha) \sim E_i$. But since $\mu(T_i) = 1$ for all i, it follows that $\mu(T) = 1$ and our theorem for $E[f]$ non-decreasing is proved. The proof of the remaining case is similar to the above.

Very little can be said about the behaviour of the error term $h_n(\alpha) - n/\log n$ in the sense of the strong law of large numbers. If we let $\psi(n) = o(1)$ and let $1 < \gamma < 2$, we can use Chebyshev's inequality to obtain

$$(42) \qquad \mu\left\{\alpha \in X\colon \left|\frac{h_k(\alpha) - \dfrac{k}{\log k}}{\dfrac{k}{\log^{\gamma} k}}\right| > \psi(k)\right\} < \frac{C}{\psi^2(k)\log^{4-2\gamma} k}.$$

If one makes the substitution $k = (1+\delta)^n$, the sum (over n) of the right hand side of (42) will converge as long as $\gamma < \frac{3}{2}$. One cannot adapt Lévy's method to this situation because the random variable $h_k(\alpha) - k/\log k$ is not monotonic. However, there are other ways to obtain strong law results from weak law results which involve making use of statistical dependency relationships between the random variables and modified versions of the

Borel Cantelli lemma. Although these methods are difficult to apply to this situation, we state as a conjecture the strongest possible result of this kind.

Conjecture 1. *If S_1 is the set of all sequences α for which*

$$h_n(\alpha) - \frac{n}{\log n} = o\left(\frac{n}{\log^{3/2-\varepsilon} n}\right); \quad \varepsilon > 0$$

then $\mu(S_1) = 1$.

On the other hand, if $\gamma > \frac{3}{2}$ in (42), we get no information from the substitution $k = (1+\delta)^n$. If we could obtain more information about the distribution of $h_n(\alpha)$ we would expect to be able to prove

Conjecture 2. *If S_2 is the set of all sequences α for which*

$$h_n(\alpha) - \frac{n}{\log n} = o\left(\frac{n}{\log^{3/2+\varepsilon} n}\right); \quad \varepsilon > 0$$

then $\mu(S_2) = 0$.

It is interesting to note that the exponent $\frac{3}{2}$ plays an important boundary role in Bateman and Diamond's work concering Beurling's generalised primes.

We will now turn our attention to a probabilistic analog of the twin primes conjecture. Although a "twin prime" is defined to be a prime p such that $p+2$ is also a prime, we will consider a "random twin" to be a pair of consecutive elements n, $n+1$. We can prove that almost all sequences in X contain infinitely many consecutive integers without making any use whatsoever of Theorem 2. However, using Theorem 2, we will obtain a strong law result which describes the asymptotic density of consecutive elements.

Theorem 4. *Let $t_n(\alpha)$ count the number of consecutive integers in the sequence α which are less than n; that is*

$$t_n(\alpha) = \sum_{\substack{k<n \\ k\in\alpha \\ k-1\in\alpha}} 1.$$

Then for almost all sequences $\alpha \in X$, $t_n(\alpha) \sim n/\log^2 n$.

Proof. The function $t_n(\alpha)$ does not lend itself to the methods detailed in Section II, since its fundamental recurrence relation analogous to (18) is

$$E[t_{n+1}] - E[t_n] = -E[x_n t_n] + \sum_{k=3}^{n-1} \frac{k-1}{n-2}\left(E[x_k^{n-k} t_k] - E[x_k^{n-k-1} t_k]\right).$$

We will employ an auxiliary function

$$(43) \qquad \hat{t}_n(\alpha) = \sum_{\substack{k<n \\ k\in\alpha \\ k-1\in\alpha \\ k\,\mathrm{odd}}} \prod_{\substack{j\in\alpha \\ j<k-1}} \left(\frac{1}{1-\dfrac{1}{j}}\right) = \sum_{\substack{k<n \\ k\in\alpha \\ k-1\in\alpha \\ k\,\mathrm{odd}}} x_{k-1}^{-1}(\alpha).$$

We note that $\hat{t}$ satisfies the recurrence relation

$$t_{n+1}(a) = t_{n-1}(a) + \frac{\chi_n(a)\chi_{n-1}(a)}{x_n}, \quad n \text{ odd},$$

where $\chi_m(a)$ is the characteristic function of m. This leads immediately to the fundamental recurrence relationships

$$E[\hat{t}_{n+1}] - E[\hat{t}_{n-1}] = \left(1 - \frac{1}{n-1}\right) E[x_{n-1}] \tag{44}$$

and

$$E[\hat{t}^2_{n+1}] - E[\hat{t}^2_{n-1}] = \left(1 - \frac{1}{n-1}\right) + 2\left(1 - \frac{1}{n-1}\right) E[\hat{t}_{n-1} x_{n-1}], \quad n \text{ odd}. \tag{45}$$

The last expectation in (45) shows that we will require a recursion for $E[\hat{t}_{n+1} x^k_{n+1}]$ for general k similar to (20).

If a is any sequence in X and B_{n+1} and B_{n-1} are the elementary sets $(a; n+1)$ and $(a; n-1)$ respectively, then we have (for n odd)

$$\mu(B_{n+1}) = \begin{cases} (1-x)^2 \mu(B_{n-1}); & n-1 \notin a,\ n \notin a, \\ (1-x) x \mu(B_{n-1}); & n-1 \notin a,\ n \in a, \\ x\big(1 - x(1 - 1/n-1)\big) \mu(B_{n-1}); & n-1 \in a,\ n \notin a, \\ x^2 (1 - 1/n-1) \mu(B_{n-1}); & n-1 \in a,\ n \in a, \end{cases}$$

and

$$x^k_{n+1}(a) \hat{t}_{n+1}(a) = \begin{cases} x^k t; & n-1 \notin a,\ n \notin a, \\ x^k \hat{t} \left(1 - \frac{1}{n}\right)^k; & n-1 \notin a,\ n \in a, \\ x^k \hat{t} \left(1 - \frac{1}{n-1}\right)^k; & n-1 \in a,\ n \notin a, \\ x^k \left(\hat{t} + \frac{1}{x}\right)\left(1 - \frac{1}{n-1}\right)^k \left(1 - \frac{1}{n}\right)^k; & n-1 \in a, n \in a, \end{cases}$$

where on the right hand side of both expressions, we have written x for $x_{n-1}(a)$ and $\hat{t}$ for $\hat{t}_{n-1}$. If we make these substitutions in the expression

$$E[x^k_{n+1} \hat{t}_{n+1}] = \sum_{B_{n+1}} \mu(B_{n+1}) x^k_{n+1}(B_{n+1}) \hat{t}_{n+1}(B_{n+1})$$

we can obtain for n odd (using the same abbreviation for x_{n-1} and $\hat{t}_{n-1}$ on the right hand side)

$$\begin{aligned} E[x^k(n+1)\hat{t}(n+1)] = {} & E[x^k \hat{t}] - 2E[x^{k+1}\hat{t}] + E[x^{k+2}\hat{t}] + \\ & + \left(1 - \frac{1}{n}\right)^k (E[x^{k+1}\hat{t}] - E[x^{k+2}\hat{t}]) + \\ & + \left(1 - \frac{1}{n-1}\right)^k E[x^{k+1}\hat{t}] - \left(1 - \frac{1}{n-1}\right)^{k+1} E[x^{k+2}\hat{t}] + \\ & + \left(1 - \frac{1}{n-1}\right)^{k+1} \left(1 - \frac{1}{n}\right)^k (E[x^{k+2}\hat{t}] + E[x^{k+1}]) \end{aligned} \tag{46}$$

which can be simplified to the estimate (n odd)

$$\begin{aligned} & E[x^k_{n+1} \hat{t}_{n+1}] - E[x^k_{n-1} \hat{t}_{n-1}] \\ & = E[x^{k+1}_{n-1}] - \frac{2k}{n} E[x^{k+1}_{n-1} \hat{t}_{n-1}] + O\left(\frac{1}{n^2} E[x^{k+1}_{n-1} \hat{t}_{n-1}]\right). \end{aligned} \tag{47}$$

Note that with $k = 0$, (46) agrees with (44).

In addition to our abbreviation for x_{n-1} and $\hat{t}_{n-1}$ on the right hand side of an equation, we will let $\Delta(E[\hat{t}x^k])$ represent the expression $E[\hat{t}_{n+1} x^k_{n+1}] - E[\hat{t}_{n-1} x^k_{n-1}]$ for n odd. These differences will be summed over the range $n = 3(2)n$ (meaning $3, 5, 7, \ldots, n-2, n$) yielding estimates for $E[t_m x^k_m]$ for m even. Use will be made of (23) through (30) modified by the assertion

$$\sum_{k=3(2)n} \frac{k^s}{M^r_k} = \tfrac{1}{2} \sum_{k=2}^{n} \frac{k^s}{M^r_k} + O(n^{s-\varepsilon})$$

which is easily obtained using (22.5). We first write

$$\Delta(E[\hat{t}x^3]) = E[x^4] - \frac{6}{n} E[x^4 \hat{t}] + O\left(\frac{1}{n^2} E[x^4 \hat{t}]\right) \tag{48}$$

which when summed yields

$$E[\hat{t}_{n+1} x^3_{n+1}] = O(nE[x^4_n]) \quad (n \text{ odd}). \tag{49}$$

Writing (47) with $k = 2$, we get

$$\Delta E[x^2 \hat{t}] = E[x^3] - \frac{4}{n} E[x^3 \hat{t}] + O\left(\frac{1}{n^2} E[x^3 \hat{t}]\right). \tag{50}$$

Summing this using (49) and a modified version of (24) yields

$$E[\hat{t}_{n+1} x^2_{n+1}] = \frac{n}{2M^3_n} + O\left(\frac{n}{M^4_n}\right). \tag{51}$$

(If we had summed 1 rather than x_{k-1}^{-1} in (43), the definition of $\hat{t}$, the first expectation on the right side of (50) would be $E[x^4]$ thus rendering the estimate (49) useless. This recursive method for solving (47) would not be possible.) Continuing in this same way, we obtain

$$E[\hat{t}_{n+1}x_{n+1}] = \frac{n}{2M_n^2} + \frac{n}{2M_n^3} + O\left(\frac{n}{M_n^4}\right). \tag{52}$$

We now sum (45) using the above estimate to get

$$E[\hat{t}_{n+1}^2] = \frac{n^2}{4M_n^2} + \frac{n^2}{2M_n^3} + O\left(\frac{n^2}{M_n^4}\right). \tag{53}$$

We can sum (44) using (16) to obtain

$$E[\hat{t}_{n+1}] = \frac{n}{2M_n} + \frac{n}{2M_n^2} + O\left(\frac{n}{M_n^4}\right) \tag{54}$$

and therefore, the variance of $\hat{t}$ is

$$\begin{aligned} V[\hat{t}_m] = E[\hat{t}_m^2] - E^2[\hat{t}_m] &= O\left(\frac{n^2}{M_m^4}\right), \quad m \text{ odd} \\ &= O\left(\frac{n^2}{\log^4 n}\right). \end{aligned} \tag{55}$$

We now observe that the hypotheses of Theorem 2 are satisfied for the random variable $\hat{t}$ if we let

$$E(n) = \frac{n}{2\log n}, \quad V(n) = \frac{n^2}{\log^4 n}, \quad \text{and} \quad R(n) = \frac{1}{2\log^2 n}.$$

Thus, we can assert that for almost all sequences $a \in X$,

$$\hat{t}_n(a) \sim \frac{n}{2\log n}. \tag{56}$$

The analysis of $\hat{t}_n$ would have produced the same results if we had counted pairs $k, k-1$, for k even rather than k odd. (56) would hold for the modified function as well and therefore if we define

$$\bar{t}_n(a) = \sum_{\substack{k<n \\ k\in a \\ k-1\in a}} x_{k-1}^{-1}(a)$$

then for almost all sequences $a \in X$, $\bar{t}_n(a) \sim n/\log n$. We now define

$$r_k(a) = \begin{cases} 1 & \text{if } k \in a \text{ and } k-1 \in a, \\ 0 & \text{otherwise.} \end{cases}$$

If a satisfies $\bar{t}_n(a) \sim n/\log n$, we use partial summation to write

$$\begin{aligned} t_n(a) = \sum_{k\le x} r_k(a) &= \sum_{k\le x} \frac{r_k(a)x_{k-1}(a)}{x_{k-1}(a)} \\ &= \frac{(1+o(1))x}{\log x} \cdot x_{n-1}(a) + \sum_{k=2}^{n-1} \frac{(1+o(1))k}{\log k}(x_{k-1} - x_k) \\ &= \frac{(1+o(1))x}{\log^2 x} + O\left(\sum_{k=2}^{n-1} \frac{k}{\log k} \cdot \frac{1}{k\log^2 k}\right) = \frac{(1+o(1))x}{\log^2 x}. \end{aligned}$$

Thus, for almost all $a \in X$, $t_n(a) \sim n/\log^2 n$ and Theorem 4 is proved.

The author would like to thank Professor David Hawkins for correcting an error which occurred in the computation of C_n and to Professor Richard Guy for help in proving Lemma 3.

References

[1] Paul T. Bateman and Harold G. Diamond, *Asymptotic distribution of Beurling's generalized prime numbers*, Studies in Number Theory, pp. 152–210, Math. Assoc. Amer., 1969.
[2] A. Beurling, *Analyse de la loi asymptotique de la distribution des nombres premiers généralisés, I*, Acta Math. 68 (1937), pp. 255–291.
[3] D. Hawkins, *The random sieve*, Math. Mag. 31 (1957–8), pp. 1–3.
[4] — *Random sieves II*, J. of Number Theory 6 (1974), pp. 1–9.
[5] Paul Lévy, *Processus Stochastiques et Mouvement Brownien*, Paris 1965.
[6] M. C. Wunderlich, *A general class of sieve generated sequences*, Acta Arith. 16 (1969), pp. 42–56; contains other references.
[7] — *The prime number theorem for random sequences*, to appear.

NOTHERN ILLINOIS UNIVERSITY
DeKalb, Illinois
UNIVERSITY OF NOTTINGHAM
Nottingham, England

Received on 22. 4. 1973 (398)