A NOTE ON PRIME NUMBERS

By S. Ducray

THE first Borel-Cantelli lemma depends upon outer measure rather than probability as such. This is utilised by forming an infinite sequence from prime-numbers and imbedding it in a sample-space. Auxiliary results for which no proof or reference is given will be found in standard textbooks on probability[1] and number-theory[2].

1. The main sequence: Transform the real half-line $x \geqslant x_0 \geqslant 2$ into $y \geqslant 0$ by $y = Li(x) - Li(x_0)$, where $Li(x)$ is the integral of $dt/\log t$ to the upper limit x. The y-line is covered by the infinite sequence of unit intervals $I_n : (n-1) \leqslant y < n$. Our *main sequence* is $\{X_i\}$ taking X_r as the number of primes p in the x-image of I_r. Thus, $X_i = 0, 1, 2, \ldots$. Define S_n and S_n^k by $S_n = X_1 + X_2 + \ldots + X_n$ and $S_n^k = (X_1)^k + (X_2)^k + \ldots + (X_n)^k$. Let A be an event defined in terms of the X's and a_n the number of times A occurs in the first n terms of the main sequence. Then we define $P^*(A) = lim.sup\ (a_n/n)$ and $^*P(A) = lim.\ inf.(a_n/n)$ as $n \to \infty$. Both P^* and $^*P$ lie between 0 and 1. The former obeys $P^*(A+B) \leqslant P^*(A) + P^*(B)$, where $A+B$ is 'the event A-or-B' in the sense of a set-union. These definitions lead to

Lemma 1.: $P^*(X - r) < ca_r/(r-1)!$; $r \geqslant 2$; a, c *constants* > 0.

This follows from a theorem of Erdös as utilised by Kosambi[3])

Lemma 2: $S_n/n \to 1$ *as* $n \to \infty$.

This is an immediate consequence of the prime-number theorem.

Lemma 3: $S_n^k/n < \infty$ *as* $n \to \infty$, *for every* $k \geqslant 2$.

This is an elementary consequence of lemma *1* and the definitions.

Lemma 4: *Each of the two inequalities*

$((1)\ldots - a\phi(n)\sqrt{n} < S_n - n < a\phi(n)\sqrt{n}$, $\phi(n) = \log\log\log n/\sqrt{\log n}$, *is false for infinitely many values of* n, *for some suitable* $a > 0$.

This is equivalent to a famous result of J. E. Littlewood. Our problem here is to determine a lower bound for $\phi(n)$ such that (*1*) is true for all values of the index n, or at least for all large n.

Lemma 5: *For any two initial points* x_0 *the difference between the corresponding* S_n *is* $0(\log n)$; *between the two* S_n^k *is* $0(\log^k n)$.

Reprinted from the Journal of the Uuiversity of Bombay, Volume XXXII, Parts 3 & 5, November 1963-March 1964.

This is obvious, as the two sequences differ by primes in a finite number of I-intervals. Here, $\log n$ could even be replaced by $\log n/\log\log n$. The result suffices to show that P^*, $^*P$ and our main theorems are independent of x_0.

LEMMA 6 : $1 > P^*(X = 0) > 0$ *and* $1 > {}^*P(X > 0) > 0$.

These follow from lemma *1* and known number-theoretic results. Given x_0 , there exists a definite procedure—the sieve of Eratosthenes —whereby X_n is uniquely determined for any n however large. There is no formula for X_n no matter what the x_0. The mere knowledge of any finite number of the X's does not suffice to determine x_0, nor to determine any other member of the sequence. Inasmuch as we always take x_0 as unspecified, $\{X_n\}$ has the unpredictability of a statistical random sample, though nothing has been said of a parent population which the word ' random ' implies. This indeterminacy conditions our approach, being vital though tacit in all that follows. Any attempt to produce a counter-example must also take into account other known properties of the primes. In particular, that the prime-number theorem holds asymptotically over y- stretches $(n, n + n^a)$, where $a < 5/8$; and even $a < 1/4$ if ' almost all n ' are meant. Not only is $X = 0$ infinitely often but the number of consecutive zeros in the main sequence exceeds $f(n)$ infinitely often, where $f \to \infty$ rather slowly, but monotonically, with n. If $X_n < b$ for all large n should happen to be true (and such a result is not proved), then $b \geqslant 2$. These ' descriptive ' properties remain in the background, being mentioned only to show what would have to be considered for pseudo-primes.

2. THE IMBEDDING SAMPLE-SPACE. Our main sequence $\{X_n\}$ admits at least one sub-sequence over which the frequency of $X = 0$ reaches a limit equal to $P^*(X = 0)$. Take such a sub-sequence and select from it a second over which a least possible limit is reached for the frenquency of $X = 1$. Then take a third subsequence over which S_n^2/n actually reaches the greatest possible limit ; and so on for maximal limits of all higher moments S_n^k/n. Should this not determine all frequencies for $X = 2, 3, 4\ldots$ completely with a corresponding final sub-sequence, choose a further set of consecutive sub-sequences such that each of these limiting frequencies is the greatest possible. Call the final limiting frequencies thus obtained for $X = r$ as F_r for $r = 0, 1, 2,\ldots$By lemma *6*, F_0 and at least one other F_i will be positive. Further :

$$((2) \ldots \Sigma F_r = 1 ;\ \Sigma r F_r = 1 ;\ \Sigma r^k F_r = M_k < \infty \text{ for all } k.)$$

These follow very simply from lemmas 1 — 3. The *imbedding sample-space* is now defined by taking independent stochastic variates X_n, each with the identical distribution defined by $P(X = r) = F_r$.

where the letter P without asterisk always denotes probability over the sample-space. A point of the sample-space, or a ' sample-sequence ' or ' sample ' is any infinite sequence of these stochastic X's . There can be no confusion because the context will always show whether the main sequence or just a sample is meant. P^* is used solely for the main sequence., P for sample-space ; *the former is not a probability.* It is easily shown that there exists a canonical mapping for the complete sample-space into $0 \leqslant t < 1$ such that there is a $1 - 1$ correspondence between the samples and the points of the unit interval. With the mapping, probability over the sample-space becomes Lebesgue measure of the corresponding set on $(0,1)$.

The law of large numbers, the central-limit theorem, the upper and the lower law of the iterated logarithm all hold over the particular imbedding sample-space constructed.

LEMMA 7 : *If S_k be the sum of (any) k consecutive X's (for the main sequence as also for the sample-space) and C a sufficiently large positive constant, then for all large k, $P^* < P$ for each of the two events $s_k - k \geqslant C\sqrt{k}$ and $s_k - k < -C\sqrt{k}$.*

Proof: The mean value of $(S_k - k)$ is zero over the main sequence by the prime-number theorem and over the sample-space by (1),. The sample-space was so constructed as to give the greatest possible scattering from the mean for any sub-sequence of the main sequence. The letter is formed only by the sieve of Eratosthenes. No prime factor of any composite number can divide any other integer for the corresponding x-distance on either side. Almost all numbers being composite, a certain number of deleting primes in the sieve are thus inactivated over any stretch of the x-line, but irregularly. Unusually many primes actually found in any stretch formed of consecutive covering intervals may mean, at worst, a slight decrease in the chances of primality in the immediate neighbourhood of that stretch. Unusally few primes so found would have the opposite effect —if any—amounting at most to a slight increase in the chances of a prime lying in nearby intervals. In either case, the tendency (*if any*) is to decrease the deviations from the mean k in sums s_k. That is, if there be any association at all between primes, it can be *compensatory on the whole but not cumulative.* For the sample-space, on the other hand, statistical independence means that the chance of an X taking on any value are not affected in any way by the values actually assumed by any number of the other X's. Comparing the two gives $P^* \leqslant P$.

LEMMA 8 : *If an infinite sequence of events A_n be defined (both over the main sequence and the sample-space), each in terms of a finite number n_k (non-decreasing) of the X's and (i) $P^*(A_n) > 0$, (ii) $P^*(A_n) \leqslant P(A_n)$*

for all large n, (iii) $\Sigma P(A_n)$ converges; then at most a finite number of the events A can occur in the original main sequence.

Proof: By the first Borel-Cantelli lemma, the result is true for the sample-space, in the sense of unit probability. That is, the probability measure is $P = 0$ for the set of samples over which an infinite number of the A's can occur. Define E_{rm} as the event $A_{r+1} + A_{r+2} \ldots + A_m$, with $m_k \leqslant n$. Then, the first m events occur in a sub-space of not more than n X's. From the convergence of $\Sigma P(A_n)$, $P(E_{rm})$ may be made arbitrarily small by taking r large enough, for all m. By hypothesis *ii*, P^* $(E_{rm} \leqslant P(E_{rm})$ while $P(E_{rm}) > 0$ by *i*. Therefore, the partial frequency with which E_{rm} occurs over almost all sample-sequences cannot ultimately be less than (say) half that which is the least upper bound for the main sequence. The reason is that, by the law of large numbers, any event which has a positive probability has the same limiting frequency over almost all points of the sample-space. If, then E_{rn} necessarily occurred over the main sequence no matter how large the index r, it occurs with comparable frequency and therefore infinitely often over almost all samples. But then, even if the limiting probabilities be zero, infinitely many of the A_r must necessarily occur over the main sequence and *a fortiori* over almost all samples. This last would contradict the findings of the Borel-Cantelli lemma set out above, i. e. $P = 1$ instead of $P = 0$. The contradiction proves our main result (which it should be noted, holds absolutely, not in the sense merely of $P^* = 0$).

3. An Application. *Definition*: The event A_r will now be said to have occurred, if for a least one k with $2^r \leqslant k < 2^{r+1}$ (3)... $s_k - k > 2^r/^2$. $2a\sqrt{Vr \log 2}$; $a > 1$, $V = \Sigma r^2 F_r - 1 > 0$)

Similarly for the event B_r with $-a$ for a and reversed inequality.

Theorem 1: *There exist two positive constants C_1, C_2 such that* 4... $-C_1\sqrt{2n \log n} < S_n - n < C_2\sqrt{2n \log n}$,
for all n, over the main sequence $\{X_n\}$ of primes in intervals I_n.

Proof: If the event A_r as defined above does not occur at all, then the right-hand inequality in (4) holds for all n with $2^r \leqslant n < 2^r$ taking $C \geqslant a.\sqrt{2V \log 2}$. The central limit theorem holds for the sample-space and gives, for all large k,

(5)... $P(s_k - k > x\sqrt{2kV}) < c.e^{-x^2}/x$; V *as in* (3).

The maximum value of the (monotonically decreasing) function on the right is attained for the least x for any range of k. Comparing (3) with (5), this least value in $2^r \leqslant k < 2^{r+1}$ is seen to be greater than $\sqrt{r \log 2}$ Now $P(A_r)$ cannot exceed the sum of the P's for each k in the range, whence

6)... $P(A_r) < c \cdot 2^r e^{-a^2 r \log 2}/a\sqrt{r \log 2} < c \cdot 2^{-(a^2-1)r}$.

therefore $\Sigma P(A_n)$ converges. Obviously, $P(A_r) > 0$ for all r while lemma 7, shows that $P^*(A_r) \leqslant P(A_r)$ for all large r estimating the P^* separately for each k in (5) and taking the sum over the given range of k. All conditions of lemma 8 being thus satisfied, at most a finite number of the A_r can occur. These last would be covered by taking a suitably large C_2.

The arguments are then repeated for the events B_r and the left-hand inequality in (4), to complete the proof. The demonstration is validated for all x_0 by lemma 5, by further adjustment if necessary of C_1, C_2,

To translate this result into number-theoretic language, let $\pi(x)$ be the number of primes $p < x$. Note that y takes on the values n in (4) and that $y \sim x/\log x$ while $S_n = \pi(x) - \pi(x_0$ for the x-value corresponding to $y = n$. Combining these with theorem 1 gives immediately

THEOREM 2 : $\pi(x) - Li(x) = 0\,(\sqrt{x})$.

The applications to number theory are too well known to be amplified here. Further refinement might be possible, given the applicability of the upper law of the iterated logarithm; the *lower* law, like the *second* Borel-Cantelli lemma upon which it rests, cannot be carried over to the main sequence.

REFERENCES

1. W. Feller : *Introduction to Probability Theory And Its Applications* Vol. I (New York, 1950), pp. 133-145 & 154-161. The proofs, given for the binomial distribution, extend readily to our case. See also J. V. Uspensky : *Introduction To Mathematical Probability* (New York, 1937), pp. 284-292.

2. K. Prachar : *Primahlverteilung* (Berlin, 1957). Also, G. H. Hardy and E. M. Wright : *Introduction To The Theory Of Numbers* (4-th ed. Oxford, 1960).

3. D. D. Kosambi : The sampling distribution of Primes ; *Proc. National Acad. Sciences* (USA) 49. 1963. 20-23 ; especially the proof of lemma 5 of that note. My debt to Prof. Kosambi is obvious.

4. S. Ducray : Normal Sequences : *J. Uni. Bombay* 31. 1962. 1-4.