STATISTICAL METHODS IN NUMBER THEORY

By D. D. Kosambi

Poona

1. Basic Results

The function $\zeta(s)$ of a complex variable $s = \sigma + it$ is defined by the Dirichlet series and the Euler product:

$$\zeta(s) = \sum \frac{1}{n^s} = \Pi \left(1 - \frac{1}{p^s}\right)^{-1}; \tag{1}$$

n, integer; p, prime; $\sigma > 1$; and by analytic continuation over the rest of the complex domain. It is known[1] that $\zeta(s)$ has no finite singularity except the simple pole $1/(s-1)$. It has no zero in $\sigma > 1$ and only the trivial zeros $s = -2, -4 \ldots$ in $\sigma < 0$. Infinitely many of its zeros lie on the vertical line[2] $s = \frac{1}{2} + it$. The Riemann hypothesis ($= RH$) is that all non-trivial zeros of $\zeta(s)$ lie on $\sigma = \frac{1}{2}$.

Let $\pi(x)$ be the number of primes $p \leqslant x$ and $li\ (x)$ the integral $\int dt/\log t$ over $2 \leqslant t \leqslant x$. Then it is further known[3] that the range of variation of $\pi(x) - li(x)$ must include $\pm x^a$ infinitely often as $x \to \infty$, where a is the greatest abscissa of any zero of $\zeta(s)$. It was proved by J. E. Littlewood[4] that there exists a number $b \geqslant 0$ such that the value of $\pi(x) - li(x)$ obeys each of the inequalities:

$$\pi(x) - li(x) > b\sqrt{x}\,\frac{\log_3 x}{\log x};$$

$$\pi(x) - li(x) < -b\sqrt{x}\,\frac{\log_3 x}{\log x}, \tag{2}$$

infinitely often as $x \to \infty$. Here, $\log_2 x = \log(\log x)$ and $\log_3 x = \log(\log_2 x)$, all logs to the base e.

Starting from the initial point $x_0 > 2$ and any fixed but arbitrary $u > 0$, the real half-line $x \geqslant x_0$ is transformed into $y \geqslant 0$ and covered by right-open intervals I_n as follows:

$$y = li(x) - li(x_0); \quad I_n : (n-1)\,u \leqslant y < nu. \tag{3}$$

The number of primes in the x-image of I_n is denoted by $\pi_n(u)$ or $\pi(x_0, u; n)$. The prime-number theorem[5]: $\pi(x) \sim li(x)$, amounts to $\Sigma \pi_n(u) \sim Nu$, summation over $1 \leqslant n \leqslant N \to \infty$. This gives:

LEMMA 1: RH *is true if and only if, for every* $\epsilon > 0$ *and some* $u > 0$:

$$\Sigma_1^N \pi(x_0;\ u;\ n) - Nu = 0(N^{\frac{1}{2}+\epsilon}). \tag{4}$$

It is essential to show that the totality of distinct sequences $\{\pi(x_0, u; n)\}$ is equivalent to the number of points on a continuous line segment. This will enable a suitable measure to be introduced. To this end, the following lemma is essential:

LEMMA 2: *There exists at least one* $u > 0$ *such that the number of distinct sample-sequences* $\{\pi_n(u)\}$ *obtained by shifting the initial point* x_0 *through a single covering interval of* y*-length* u *can be put into a* $1-1$ *correspondence with the points of* $0 \leqslant t < 1$.

Proof.—Suppose that, for some given u and x_0, the same sequence $\{\pi_n(u)\}$ is obtained when the initial point is shifted to the right through a y-distance w. It would then follow that the number of primes gained by any interval at the right is precisely equal to that lost at the left during the shift. Therefore, every w-interval, separated by the y-distance $u - w$ from the next on either side must contain the same number of primes. Known separation theorems[6] by P. Erdös say that there exist infinitely many gaps between consecutive primes, larger than any preassigned y-length. Hence these w-intervals must be totally void of any primes.

The results of G. Ricci[7] show, on the other hand, that there exist sub-sequences of primes such that the y-distances between consecutive primes are dense over some non-zero interval $(1-\alpha,\ 1+\beta)$. If $\alpha = 1$, then take any $u < 1 + \beta$. Otherwise, take an integer k so large that $(1-\alpha)/k$ is less than $\alpha + \beta$, and take $(1-\alpha)/k = u$. In either case, the Ricci density theorem shows that w must vanish. Thus, for the chosen u (and there are infinitely many such choices, obviously), there must be as many distinct sequences as points of $0 \leqslant t < u$; this can be projected upon the unit interval $0 \leqslant t < 1$, to complete the proof of the theorem (which holds in fact for all $u > 0$).

LEMMA 3: *If* $M = M(z)$ *be the product of all primes* $p \leqslant z$ *and* γ *is Euler's constant, then the number of integers relatively prime to* M *in any range* $A \leqslant n < A + R$ *is asymptotic to* $Re^{-\gamma}/\log z$ *as* $z \to \infty$,

provided $R/\log z$ *is large compared to* $2^{\pi(z)}$, *where* $\pi(z)$ *is the number of primes* $p \leqslant z$.

Proof.—This is essentially the form in which the Sieve of Eratosthenes is to be used. The integers prime to M are cyclically arranged modulo M with symmetry about the middle of any cycle kM to $(k+1)M$. If $A = kM + 1$, the number out of the R consecutive integers, not divisible by any prime $p \leqslant z$ is given by:

$$R - \left[\frac{R}{p_i}\right] + \left[\frac{R}{p_i p_j}\right] - \left[\frac{R}{p_i p_j p_k}\right] + \cdots; \quad p_i, p_j, p_k \neq . \qquad (5)$$

The square brackets denote the largest positive integer in the enclosed quotient, or zero. The primes are to run through the complete set $p \leqslant z$. Since no remainder can be as great as unity, the difference when the brackets are removed will not exceed $\frac{1}{2}(1+1)^{\pi(z)}$ in absolute value. For any A, we can regard the result as the sum of difference of two expressions as in (5). The asymptotic value of $R\Pi(1 - 1/p)$, which is the value of (5) with the brackets removed is $Re^{-\gamma}/\log z$ by the classic theorem of Mertens.[8]

2. Lemmas on Measure

Definitions.—A proper frequency distribution is furnished by a set of real numbers $f_i > 0$ such that $\Sigma f_r = 1$. If $A_0, A_1, A_2, \cdots$ be an indexed set of distinct *attributes*, an infinite sequence thereof $A_i A_j A_k \cdots$ (not necessarily all distinct) represents a *sample*, or point in *sample-space*. A sequence $\{A_r\}$ wherein the limiting frequency with which a particular A_n occurs is, for every n, the f_n above has that distribution. By *probability* is meant a measure function obeying the usual postulates, defined over the whole sample-space or over a sub-set thereof, such that the total measure of the universe of definition is unity. The probability measure of an *event* (sub-set of sample-space) is indicated by the letter P. The n-th term of a sequence has the designation X_n and $P(X_n = A_j)$ is the probability measure, if it exists, of the set of sample-points where A_j appears as the n-th term of the corresponding sequence. The joint probability of a *compound* event is similarly defined, *e.g.*, $P(X_i = A_j;\ X_r = A_k \ldots)$. A sample-sequence is *normal* if every finite combination $A_i A_j A_k \ldots$ occurs with frequency equal to the product of the individual component frequencies. Correspondingly, the events $X_i = A_j,\ X_r = A_k \ldots$ are said to be *independent* in probability if for any number of such events the compound probability is the product of the component individual probabilities.

LEMMA 4: *Given a set of attributes* $A_0, A_1, A_2, \cdots$ *and a corresponding proper frequency distribution. Then there exists a mapping whereby*: (1) *The totality of sample-sequences is mapped in a* $1-1$ *manner onto the right-open unit interval* $0 \leqslant t < 1$. (2) *The Lebesgue measure on the map is equivalent to probability measure over the sample space.* (3) *Almost all sample-sequences are normal with the given basic frequency distribution and all the events* $X_i = A_j$ *are independent in probability.*

Proof.—The actual map is constructed as follows. Divide (0, 1) into right-open sub-intervals by marking off successive points $t_0 = f_0$), $t_1 = f_0 + f_1, \cdots t_i = f_i + t_{i-1}, \cdots$ Then subdivide the sub-intervals $(0, t_0)$, $(t_0, t_1), \cdots$ in the same manner, *each* in proportion to its total length. And so on, step by step. For the sequence $A_i A_j A_k, \ldots$, take first the sub-interval immediately to the left of t_i in the first subdivision. Then in the next subdivision of this selected interval, that to the left of the point marked off with the subscript j; and so on, taking the next stage of subdivision for each successive subscript. The sequence of nesting intervals obviously converges to a single point in (0, 1). Conversely, to each such point there corresponds just one sequence of subscripts, provided a suitable convention is made (to avoid duplication) about sequences terminating in an infinite succession of zeros or of the final index r when the total number of frequencies is finite and equal to $r + 1$. The properties listed follow obviously, with this mapping.

The well-known theorem of Borel[9]: *almost every number in* (0, 1) *is normal in a 'decimal' expansion to any base* becomes a special case of this lemma when the number of attributes are finite, with equal frequencies. The proof, for finite or infinitely many basic frequencies, may be derived from the law of large numbers[10] in probability theory.

LEMMA 5: *Given a sample-sapce where the basic frequencies have a Poisson distribution with parameter* u, *the sample-sequences are normal, and the attribute* A_r *assigned the numerical value* r. *Then almost all points of the sample-space obey the inequalities*:

$$-(1+\epsilon)\sqrt{2Nu\log_2 Nu} < \Sigma_1^N (X_i - u) < (1+\epsilon)\sqrt{2Nu\log_2 Nu}, \tag{6}$$

with at most a finite number of exceptions as $N \to \infty$, *for every* $\epsilon > 0$.

Proof.—This is the *upper* law of the iterated logarithm, abbreviated *ULIL*. The Poisson distribution has $f_i = e^{-u}u^i/i!$. The standard

proof[11] for binomial distributions extends immediately to the Poisson, hence need not be repeated here. The canonical mapping of Lemma 4 is to be used.

LEMMA 6: ULIL *holds with unit probability for all* $\epsilon > \lambda - 1 > 0$ *if the* X_t *of lemma 5 have the Poisson distribution with parameter* u *and a joint distribution such that*: (1) *The sum of any finite number k of consecutive* X_r *has the Poisson distribution with parameter* ku. (2) *The probability that* $|\Sigma_1^k (X_r - u)| > \lambda \sqrt{2Nu \log_2 Nu}$ *for some* $\lambda > 1$ *and at least one* $k \leqslant N$ *does not exceed the corresponding probability when the* X_r *have distributions independent in probability; for all large* N.

Proof.—Lemma 5 does not depend upon any particular mapping; nor is independence necessary, though it suffices. The *first* Borel-Cantelli lemma[12] upon which *ULIL* depends does not require independence. The two conditions given here suffice for the text-book proof of Lemma 5 cited,[11] as may be verified by inspection.

3. APPLICATIONS

In what follows, only the sample-sequences $\{\pi(x_0, u; n)\}$ are considered. The attribute A_k will be taken to have presented itself whenever a member of such a sequence has the value k. Again, X_r is simply the numerical value of the r-th member of such a sequence. Then we have:

LEMMA 7: *The sequences* $\{\pi_n(u)\}$ *have the Poisson frequency distribution with* $f_r = e^{-u}u^r/r!$, *in the sense of unit probability measure.*

Proof.—This follows from known[13] results, and could be proved again from the following considerations: As the number of trials (integers tested per covering interval) increases, the probability of the event (of a number being prime) tends to zero, but nevertheless the expectation (primes 'expected' per interval) tends to u; and the probability is unaffected by the results of any number of previous trials, or at worst, the change in the probability is an infinitesimal of higher order than P itself. This last point is proved in the next Lemma; the rest are obvious.

There now arise three possibilities:

(A) The sequences $\{\pi_n(u)\}$ are *independent* in probability, in the sense that the actual values of any finite number of X's will not determine x_0, nor affect the probability for a given value of any other X_r to occur. In that case, *ULIL* of Lemma 5 and therefore Lemma 1 would

hold, hence *RH* is true. In addition, the *lower* law of the iterated logarithm would also apply, which would enable the Littlewood inequalities (2) to be improved with the bounds replaced by $\pm (1 - \epsilon) \sqrt{2 x \log x \log_2 x}$.

(B) The sequences may not be independent, but the effect of any dependence upon the sums of consecutive members may be *compensatory*. That is, deviations in the sums from expectation might be no greater (in probability) than in case A. It suffices if the probability measure of the set $| \Sigma \pi_n (u) - Nu | > a$ (summation over indices 1 to N) does not exceed that in the case A. Then Lemma 5 and *ULIL* could still hold, but (2) cannot be improved and the Littlewood inequalities might be the best possible.

(C) The effect of dependence (if any) might be *cumulative*. That is, the occurrence of an excess from expectation in either direction, for sums of consecutive $\pi_n (u)$ might imply a similar excess in the same direction somewhere else in the same sequence (positive autocorrelation). In this case, *ULIL* need not hold, nor *RH*.

The sieve of Eratosthenes, as will be seen, excludes C.

LEMMA 8: *The terms of a sequence* $\{\pi_n (u)\}$ *are asymptotically independent in probability; moreover, the effect (if any) upon sums of consecutive terms of any deviation from independence cannot be cumulative, but at most compensatory.*

Proof.—In the discussion that follows, consider only such covering intervals as lie in the range $(x/2, x)$. This suffices because: (*i*) The prime-number theorem (and hence also the Poisson distribution) is asymptotically valid over such lengths of the real half-line; in fact, even over much smaller ranges, $(x, x + x^a)$ if $a > 38/61$, as is known.[14] (*ii*) The proof and applications of *ULIL* may be carried through with successive ranges of order (Ab^k, Ab^{k+1}), with any fixed $b > 1$ and $k = 1, 2, 3, \ldots$, so that there is no loss of generality involved. In the discussion, however, x is only to be regarded as a large background parameter whose sole use is to estimate the relative magnitudes of various arithmetic functions that appear. If x were specified exactly, there would be no question of probability, as everything would be exactly known.

In the sieve of Eratosthenes, the multiples of 2, 3, 5, ... are successively deleted; at each stage, the smallest number left is the next prime to be used in deletion. This way, every prime and only primes are obtained, as a succession of smallest survivors of the deletions

Every integer is deleted by its *smallest* prime factor, and once deleted, so remains regardless of how many other primes divide it. If this division were independent (in the sense of probability theory), there would be nothing left to prove, and alternative A above would be the only one left. Lemma 3, however, says that primes $p \leqslant h$, where h is the length of an interval I_n hence $h \approx u \log x$ have multiples in every I_r and act independently (with probability $1/p$ each) over the given range, or indeed any range of order not less than $x^{c/\log_2 x}$. Beyond this, it is not possible to go. The larger the primes, the less chance of several of them dividing an integer in the range. If independence in division were present, the Mertens theorem[3] would have given us for the prime number theorem $2e^{-\gamma}x/\log x$, instead of $x/\log x$. This is taken by some to show that "probability methods do not apply in prime-number theory", but is in fact irrelevant. The independence in probability of the number of primes in various $\pi_n(u)$, specified only by the index, not with *a priori* reference to the number of primes contained nor by knowledge of the initial point x_0, could still be a result of the sieve. The crucial question is: Given that a certain number of primes has actually occurred in a given stretch (*i.e.*, a given number of consecutive I_n), what can then be said of the chances of primality anywhere else as affected or unaffected by this occurrence?

Directly, we are concerned only with deletions by primes $p \leqslant \sqrt{x}$ The small primes act independently over the range, by Lemma 3, as noted above. The only effect that the occurrence of a composite number can have is that its prime factors will not operate in the immediate neighbourhood; for every such inoperative prime, the probability will be *locally* enhanced by a factor $1/(1 - 1/p)$. But a certain number of such dividing primes must become inoperative on the average over any stretch, while the probability for primality and expectation are always given overall by the prime-number theorem. This means that unusually many inoperative primes may cause a local enhancement of the probability for primality; unusually few, a lowering of the probability for primality. Otherwise, nothing can be said. For deleting primes between h and $x^{1/\log_2 x}$, the inoperative primes must be greater than $x^{1/\log_2 x}$ and the local factor can be calculated by packing the maximum possible number of primes lost as close to $x^{1/\log_2 x}$ from above as possible. The extreme factors are thus easily shown to be bracketed by $(1 \pm \log_2^2 x/\log x)$. For deleting primes not exceeding $x^{1/k}$, $k > 2$ fixed, the loss or gain will be not greater in either direction than $(1 \pm k \log_2 x/\log x)$. In each case, the sign makes the extreme factors compensatory, while smaller factors in any case

cannot be cumulative. Finally, for stretches of length $\sqrt{x}$ or more all the deleting primes have multiples. Unusually many deletions means unusually many factors higher than $\sqrt{x}$; again, the tendency cannot be cumulative, and the foregoing shows that the probability is changed by very little; asymptotically, not changed at all.

THEOREM 1: (RH) *No sample-sequence* $\{\pi(x_0, u; n)\}$ *can lie within the exceptional set of probability measure zero with respect to the* ULIL *of Lemma* 6, *for any* $\epsilon > 0$. *Whence all non-trivial zeros of* $\zeta(s)$ *lie on the vertical line* $s = \frac{1}{2} + it$.

Proof.—Starting with any x_0 and some fixed u derived from Lemma 2, map all sequences with initial points in I_1 onto (0, 1). The entire coset to be obtained by the displacement of any initial point in I_1 by an integral number of intervals in either direction is also mapped upon the same point of (0, 1). All members of a coset have clearly the same limiting-frequency properties. Probability measure is now taken as Lebesgue measure over the coset map on (0, 1). The probability can be calculated as the Lebesgue integral of the corresponding frequency. Thus, the basic distribution is Poissonian with parameter u, by Lemma 7. The distribution for k consecutive covering intervals amounts to that with covering intervals of length ku, and is therefore Poissonian with parameter ku. As for the condition 2 of Lemma 6, we note that the expectation per stretch of length ku on the y-line is ku primes, and that, for large x, it is physically impossible for the deviation from this expectation to be as much as $\sqrt{2Nu.\log_2 Nu}$, if k is small enough, *e.g.*, the total stretch covered by ku consecutive intervals does not exceed $\sqrt{x \log x}$. Here, the P is zero hence less than with total independence, as one would expect from the compensatory effect found above. Whatever the extreme actually found in stretches of this order, one can repeat the arguments of lemma 8. Thus, condition 2, of Lemma 6 is also satisfied by virtue of the non-cumulative effect of the sieve, and for every $\epsilon > 0$. Therefore, *ULIL* applies to almost all sample-sequences $\{\pi_n(u)\}$. For large N, the partial sums of the first N terms of any two sample sequences whose initial points lie within the y-distance u of each other cannot differ by more than $u \log Nu$. Therefore, either all the sample sequences satisfy *ULIL*, or none. The latter case is excluded, as the measure of the exceptional set would then have to be unity instead of zero. This proves the theorem and the Riemann hypothesis.

THEOREM 2: *The non-trivial zeros of all Dirichlet* L-*functions likewise lie on the vertical line* $s = \frac{1}{2} + it$.

This is the extended *RH* or the Piltz conjecture. The result is merely stated without proof, because the same methods and arguments as above suffice. The consequences of these two theorems are given in books on specialized function theory[2] and advanced number theory.[3] Improvement of the inequalities (2) by the present methods would depend upon *LLIL* and hence the *second* Borel-Cantelli lemma, which requires independence in probability.

The Poisson distribution of Lemma 2 allows many new results to be obtained directly. For example, *gaps of* y-*length* t *or more between consecutive primes have the distribution function* e^{-t}. However, it should be noted that the Poisson distribution is not essential for *RH*, which can be proved without any distribution at all, merely by taking the Poissonian as a bounding distribution for estimating the deviations of sums from expectation. Also, the Poisson distribution would not follow directly, granted *RH*. In other words, Lemma 8 is more important than Lemma 7.

Counter-examples of a fairly complicated nature could be produced which do not affect Lemma 7 nor the inequalities (2) but for which *RH* is false. These are formed by adding pseudo-primes and by striking out (in suitable stretches) sufficiently 'thin' sequences of the primes. Such counter-examples do not affect our arguments because such changes in the series of prime numbers within the positive integers will block sieve deletion, invalidate the Euler product, and destroy unique factorization—all of which are essential to *RH* (as they are to our present arguments).

The result also indicates that the zeros of $\zeta(s)$ on the vertical line $\sigma = \frac{1}{2}$ should have a distribution of their own, presumably also the Poisson distribution. The proper transformation here must replace the co-ordinate t on the vertical line by the integral $\int \log t dt$ to the upper limit $T/2\pi$. The results will be considered elsewhere.

REFERENCES

1. Titchmarsh, E. C. .. *The Theory of Functions*, Oxford, 1932, p. 152.
2. ———— .. *The Zeta-Function of Riemann*, Oxford, 1951, pp. 214–22; Chapters xiii-xiv give the consequences of *RH*.
3. Prachar, K. .. *Primzahlverteilung*, Berlin, 1957, pp. 247–55.
4. Littlewood, J. E. .. "Sur la distribution des nombres premiers," *Comptes Rendus*, Paris, 1914, **158**, 1869–72 and note 3 above.

5. Prachar .. *Loc. cit.*, Chapter ii.

6. ——— .. *Ibid.*, p. 157, *et seq.*

7. Ricci, G. .. *Rendiconti, Atti. Acad. Naz. Lincei*, 1954–55, **8**, 192–96 and 347–51. Prof. P. Erdös was kind enough to inform me that the Ricci statement must be emended to: 'the set of cluster points for the y-distance between consecutive primes is of positive measure'. Even this suffices to prove lemma 2, except that in this case, not every displacement however small need lead to a different sequence throughout the u-interval. This will be replaced by a set of sub-intervals of the u-interval, whose total length remains positive, and may then be mapped upon (0, 1).

8. Hardy, G. H. and Wright, E. M. .. *An Introduction to the Theory of Numbers*, II Edition, Oxford, 1945, pp. 349–54, Theorem 430.

9. ——— .. *Ibid.*, Sections 9.12 and 9.13.

10. Feller, W. .. *Introduction to Probability Theory and Its Applications*, New York, 1950, **1**, pp. 161–63; proof for finite base only.

11. ——— .. *Ibid.*, pp. 158–59, *cf.* Feller in *Trans. Am. Math. Soc.*, 1943, **54**, 373–402.

12. ——— .. *Ibid.*, p. 154.

13. Kosambi, D. D. .. "The sampling distribution of primes," *Proc. Nat. Acad. Sci.* (U.S.A.), 1963, **49**, 20–23.

14. Ingham, A. E. .. *Quart. J. Math.* (Oxford), 1937, **8**, 255–66.