1021

KOSAMBI, D. D.

"THE RIEMANN HYPOTHESIS"

1958

[with TLS 8.22.56]

The Riemann Hypothesis

by

D.D.Kosambi (Poona)

The basic result of this paper, from which important conclusions follow, is the proof that the series

$$(1)\ldots \qquad \sum_{n \leq x} 1/n^{\sigma} \; - \; \sum_{p \leq x} \log p/p^{\sigma} \quad , \quad x \longrightarrow \infty$$

converges for every real $\sigma > \frac{1}{2}$. Here, n runs through the positive integers and p the primes, in natural order. The convergence of (1) for $\sigma \geq 1+0$ is known, but not for $\frac{1}{2} < \sigma \leq 1$. The following method is adopted for our proof. Terms of the series (1) are grouped together for n and p in irregular, non-overlapping consecutive intervals d_{ν}. With a suitably chosen, not necessarily integral, real number x_{ν} in d_{ν}, the original series (1) is then replaced by

$$(2)\ldots \qquad \sum_{\nu} (d - \log x)/x^{\sigma} \quad , \qquad \sigma > \frac{1}{2} \quad ,$$

where the subscript ν has been omitted for convenience from d and x. Differences between the partial sums of (1) and (2) arise, namely : Due to the grouping, and the substitution of x_{ν} for n and p; to the irregularities in the number of primes contained in any d-interval; and lastly because the partial sums of (1) and (2) will not in general stop at the same term.

If, however, a covering set of intervals $\{d_{\nu}\}$ exists such that (2) and the various series and sequences (finite in number) arising from the above differences all converge simultaneously, then clearly the series (1) converges.

In what follows, $\pi(d)$ indicates the number of primes in the interval d. The prime number theorem is taken for granted in the form: $\pi(0,x) \sim \underline{\mathrm{li}}\ \underline{x} \sim x/\log x$. P denotes a probability, E the expectation (mean) and V the variance (dispersion) in the sense of probability theory. For stochastic X and scalar λ, we always have $E(\lambda X) = \lambda E(X)$ and $V(\lambda X) = \lambda^2 V(X)$. By a variate is meant a stochastic variable, i.e. one that has a probability distribution. The following result of A.Kolmogoroff[1] is fundamental:

Lemma K: The stochastic series of independent variates $\sum u_n$ converges with P=1 if there exists another set of independent variates $\{v_n\}$ such that the series $\sum P(u_n \neq v_n)$, $\sum E(v_n)$, and $\sum V(v_n)$ all converge; otherwise the convergence-probability of $\sum u_n$ is zero.

The use of this theorem in the sequel does not mean that (1) converges with unit probability, for (1) is not a stochastic series. The utility of lemma K lies in showing the existence of a suitable choice of d-intervals. That is a stochastic mechanism of selection may be set up for $\{d_\nu\}$, $\nu = 1, 2, \ldots$ so that (2) and the series and sequences of its differences with (1) all converge with P=1. Therefore, at least one infinite sequence of covering intervals $\{d_\nu\}$ must exist giving the simultaneous convergence of all these and hence the series (1) converges. The existence theorem does not actually construct a specific set of $\{d\}$, but the logic involved is completely rigorous, having as its basis the fact that a set of positive measure cannot be empty. The proof is not heuristic, as it would have been had d been taken as the interval between consecutive primes, or the prime numbers been otherwise treated as a stochastic sequence because of their irregularity. An important feature of Lemma K is that P=0 or 1, no other value so that $P>0$ implies P=1 immediately. Extensions to series of variates not completely independent, or where the variances themselves form another stochastic series, can be used.

1. The real line $2 \leq x \leq \infty$, on which the integers and primes are marked off, is transformed into $0 \leq y \leq \infty$ by $y = \underline{li}\ x = \int dx/\log x$. Then, to an interval $(a, a+\Delta)$ on the y-line corresponds a unique interval d on the x-line and conversely, with $\Delta = d/\log x$; where x is chosen as that number (not necessarily an integer) lying in d, which makes this relationship hold. The mean-value theorem shows the existence of such an x, which lies properly within d. The intervals may be taken to include the left-hand end point, but not the right. An arbitrarily large initial portion of either line may be ignored in discussion of the convergence problem.

~~It remains to specify the distribution of the stochastic but independent~~ . Each length Δ_ν, $\nu = 1, 2 \cdots$ is taken to have ~~has~~ the identical distribution, namely, the uniform distribution over (0,2). Being open on the right, consecutive marking off intervals of the lengths Δ_ν without gaps furnishes a complete non-overlapping covering of the y-line. That is, the length is a stochastic variable equivalent to the ν th independent selection from a uniform distribution; the position of the interval of length is uniquely determined by the covering process. Hence the d-intervals that correspond by the inverse li x transformation give a stochastic covering of the x-line, for each such random sample of the Δ's. The number $x_\nu \subset d_\nu$ has been specified above. For the lengths Δ_ν, we have for every ν and any positive integer k:

$$(1.1)\ldots\ E(\Delta) = 1\ ;\ V(\Delta) = 1/3\ ;\ E(\overline{\Delta - 1}^{2k}) = 1/(2k+1)\ ;\ E(\overline{\Delta - 1}^{2k+1}) = 0\,.$$

With lemma K, this leads at once to :

Theorem 1.1: The stochastic series $\sum(\Delta-1)^k \log^r \nu/\nu^{\sigma}$ converges for every odd positive integer k, any r, and all $\sigma > \frac{1}{2}$, with P=1; the series $\sum \Delta^k \log^r \nu/\nu^{1+\sigma}$ for all $k>0$, all r, and $\sigma > 0$, ~~also with P=1~~.

Proof: This follows immediately from (1.1), because the means for the first series are all zero, while the variances for each k are a fixed multiple of $\log^{2r}\nu/\nu^{2\sigma}$, and the sum of such terms converges for $2\sigma>1$. Similarly, for the second series, where $E(\Delta^k) = 2^k/(k+1)$, and $V(\Delta^k) = k^2 2^{2k}/(k+1)(2k+1)$, but no probability argument is necessary, because $0 \leq \Delta \leq 2$.

The prime number theorem allows an estimate to be made of the number of primes in d:

Lemma 1.1: $E(\pi(d)) = 1$, and $V(\pi(d)) = O(d+1)$ for any given large index ν.

Proof: The first statement follows immediately from the prime number theorem. The number of primes is asymptotic to $y_\nu = \Delta_1 + \ldots + \Delta_\nu$; and divided by ν, this tends in probability to the limit $E(\pi(d))$. But the same limit in probability is also $E(\Delta) = 1$, because of the parent Δ-distribution. For the second part, note that ~~we have to keep the index fixed. Then~~ the mean value is asymptotic to d/log x. We then apply the probability result :

Lemma 1.2: If a non-negative variate X have a fixed upper bound h and expectation h/a, its variance cannot exceed $h^2/a - (h/a)^2$.

Proof: The variance $V(X) = E(X^2) - E^2(X)$, by definition, and $E(X) = h/a$ by hypothesis. The greatest value of $E(X^2)$ can be realised only when X is concentrated at the two extreme values 0 and h. The probabilities must be $(1-1/a)$ and $1/a$ respectively to give the mean <u>h/a</u>, so that $E(X^2)$ cannot exceed h^2/a.

The number of integers, and <u>a fortiori</u> that of primes in <u>d</u> cannot exceed <u>$d+1$</u>, while $d/\log x = \Delta$ by definition, and cannot exceed <u>2</u>; these estimates substituted in the findings of lemma 1.2 complete the proof of lemma 1.1.

<u>2.</u> Theorem 1.1 can be paraphrased by putting $d/\log x$ for Δ, but its usefulness develops only when it is shown that <u>ν</u> may be replaced by <u>x_ν</u> in the coefficients of the random series:

<u>Theorem 2.1</u>: <u>The stochastic series</u> $\sum (d/\log x - 1)\log^r x/x^\sigma$ <u>and</u> $\sum d^k \log^r x/x^{1+\sigma}$ <u>converge for all r and k with</u> $P=1$, <u>when</u> $\sigma > \frac{1}{2}$ <u>and</u> $\sigma > 0$ <u>respectively</u>.

Proof: It is easy to see from the definitions that for large <u>x</u> and <u>ν</u>,

$$(2.1)\ldots \qquad \frac{x_\nu}{\log x_\nu} < y_\nu < x_{\nu+1} \quad .$$

Inasmuch as y_ν is the sum of an unbiassed random sample of variates with the same parent distribution which is symmetric with suitably bounded even order central moments (cf. formula 1.1), the estimates of S. Bernstein[2] apply. For the given uniform distribution of Δ, these can be stated as

<u>Lemma 2.1</u>: <u>The probability is less than $\exp(-t^2)$ for each of the inequalities to hold for all large ν</u>:

$$(2.2)\ldots \qquad y_\nu > \nu + t\sqrt{2\nu/3} \quad ; \quad y_\nu < \nu - t\sqrt{2\nu/3} \ .$$

Thus, taking $t = \sqrt{\log \nu}$, P is less than $\nu^{-3/2}$ for each of the inequalities :

$$(2.3)\ldots \qquad x_\nu > 2\nu \log \nu \quad ; \quad x_\nu < \nu/2 \ ,$$

to hold for all large ν, where no attempt has been made to get the best possible constants.

From this, it follows that, for a suitably chosen positive constant c, the infinite product $\prod P(x^{-\sigma} \leq c\nu^{-\sigma})$ converges, $\sigma > \frac{1}{2}$. But for every $\varepsilon > 0$, however small, any k, and large x_ν, $\log^k x/x^\varepsilon \to 0$, monotonically. Therefore, in that probability region, the positive monotonically decreasing coefficients $x^{-\sigma}.x^{-\varepsilon}.\log^k x/x^\varepsilon$ are less than $c\nu^{-\sigma}$ by a monotonic, bounded, factor of proportionality. The results for the convergence of infinite series derived from Abel's lemma[3)] are applicable, so that the series in theorem 2.1 converge with a positive probability for the exponent $\sigma' = \sigma + 2\varepsilon$, for every $\sigma > \frac{1}{2}$ and every $\varepsilon > 0$ whenever the corresponding series in theorem 1.1 converge. The compound probability is the product of the two convergence-probabilities. But the latter of the two probabilities is unity, and the former is given by the infinite product above, from which any finite number of terms may be deleted, without affecting convergence. Such omission brings the probability arbitrarily close to P=1-0, but the probability of convergence is fixed, hence must be unity; so also the compound probability. Finally, the exponent $\sigma' = \sigma + 2\varepsilon$ is arbitrarily close to $\frac{1}{2} + 0$ which completes the proof of the theorem.

The reason for not using the estimates $E(d^k) = O(\log^k x)$ directly in the proof is that the terms of the series are no longer strictly independent, and the joint distributions would also have to be worked out term by term. Thus, the use of theorem 1.1 as a first stage and the calculation of a probability for the bounded monotonic factor is essential.

It follows that for stochastic covering intervals, the series (2) also converges with unit probability. The differences between this series and (1) may be estimated piecemeal as dominated by or ~~comparable~~ equivalent to

$$(2.4)\ldots \sum_\nu \frac{d(d+1)}{x^{\sigma+1}} \;;\; \sum_\nu \frac{\pi(d)(\sigma \log x - 1)}{x^{\sigma+1}} \;;\; \sum_\nu \frac{\{\pi(d) - 1\}\log x}{x^{\sigma}} \;;$$

$$(d+1)/x_\nu^\sigma \;;\; \pi(d).\log x/x_\nu^\sigma \;.$$

The first of these arises from grouping all the terms $1/n^{\sigma}$, $n \subset d_{\nu}$, as a single term d/x_{ν}^{σ}; the series converges with P=1, by theorem 2.1. The second is due similarly to differences from each ~~contraction~~ of the terms $\log p/p^{\sigma}$, $p \subset d$, ~~into~~ and $\log x/x^{\sigma}$; inasmuch as $\pi(d) \leq d+1$, and the coefficients are positive also, theorem 2.1 again gives P=1 for the convergence. The third series originates from taking just one of these grouped prime terms into (2) as $\log x/x^{\sigma}$, and ignoring the rest. Here, $\pi(d)-1$ is a stochastic function of Δ with mean zero, and variance $O(d+1)$, by lemma 1.1. The same reasoning that carried theorem 1.1 into theorem 2.1 therefore applies, ~~and~~ though lemma K has to be applied a second time to the series of variances. The final probability of convergence is still unity. The two sequences at the end of (2.4) are estimates dominating the difference between the partial sums of (1) and those of (2); the results for stochastic sequences corresponding to those for series in lemma K again give unit probability of convergence. Moreover, the compound probability for simultaneous convergence of (2) and all members of (2.4) is still unity, though it would suffice if it were only positive. Thus, we conclude that at least one specific choice of d-intervals exists for which (2) and (2.4) all converge simultaneously. This proves

Theorem 2.2: The series (1) converges for all real $\sigma > \frac{1}{2}$.

Although $\sigma > \frac{1}{2}$ is the best possible exponent for this method of proof it does not necessarily follow from the preceding that the series (1) diverges for $\sigma \leq \frac{1}{2}$. For, even zero probability might still permit the existence of at least one suitable choice of covering intervals. For P=1, almost every choice will suit.

Our methods lead obviously to a proof of

Theorem 2.3: If $f(x) > 0$ be a monotonically decreasing function of the positive real variable x, with a continuous first derivative $f'(x)$, such that the series $\sum f^2(n).\log^2 n$ and $\sum f'(n).\log^2 n$ both converge, then the series

$$(2.5)\ldots \qquad \sum_{n \leq x} f(n) - \sum_{p \leq x} f(p).\log p\ , \qquad x \to \infty$$

converges.

3. The function $\zeta(s)$ is defined for a complex variable $s = \sigma + it$ with σ, t real, for the half-plane $\sigma > 1$ by

$$(3.1)\ldots \qquad \zeta(s) = \sum_{1}^{\infty} 1/n^s = \prod_{p} 1/(1-p^{-s})\ .$$

Both the series and the infinite product converge for the half-plane $\sigma > 1$. The function $\zeta(s)$ defined by the series and its analytic continuation has no singularity in the entire finite plane except for the simple pole with unit residue, $1/(s-1)$.

The zeta-function obeys the functional equation[4]:

$$(3.2)\ldots \qquad \zeta(1-s) = 2^{1-s}\pi^{-s}(\cos \pi s/2)\Gamma(s)\,\zeta(s).$$

The Riemann hypothesis (RH) is the conjecture that all zeros of (s) not on the negative real axis lie on the vertical line $\sigma = \frac{1}{2}$. It is easily shown, directly from the series, that no zero can lie beyond $\sigma > 1$. Using the functional equation, it would suffice to prove RH if it could be shown that no zero lies in the critical strip $\frac{1}{2} < \sigma \leq 1$. To this end, we use a classical lemma of function theory: *Any singularity of an analytic $F(z)$, except isolated simple poles with unit residue, and any zero of $F(z)$ is a singularity of $F(z) + F'(z)/F(z)$.* The simple poles $1/(z-a)$ cancel out, but zeros of $F(z)$ appear as first degree poles in the second term, the logarithmic derivative. For $F(s) = \zeta(s)$, the fact that $\zeta(s)$ has no finite singularity other than the pole $1/(s-1)$ would mean that the singularities of $\zeta'(s)/\zeta(s) + \zeta(s)$ must be due only to the zeros of $\zeta(s)$.

Formally, differentiation of the logarithm of the infinite product in (3.1) gives, using the series expansion $\log(1-x) = -x - x^2/2 - x^3/3 - \ldots$:

$$(3.3)\ldots \quad -\zeta'(s)/\zeta(s) = \sum_p \log p/p^s + \sum_p \log p/p^{2s} + \sum_p \log p/p^{3s} + \ldots$$

The expansion is valid for $\sigma > 1$. For $\frac{1}{2} < \sigma$, all the series on the right except the first are together dominated by $2\sum \log n/n^2 = -2\zeta'(2\sigma)$. Therefore, the discussion by means of $\zeta(s) + \zeta'(s)/\zeta(s)$ reduces to showing that the series

$$(3.4)\ldots \quad \sum_{n \leq x} 1/n^s - \sum_{p \leq x} \log p/p^s \quad , \quad x \rightarrow \infty \ , \quad s = \sigma + it \ ,$$

converges for all $\sigma > \frac{1}{2}$. But we have already shown that (1), which is the form assumed by (3.4) on the real axis, converges for all $\sigma > \frac{1}{2}$. Therefore, the Dirichlet series (3.4) converges throughout the open half--plane to the right of $\sigma = \frac{1}{2}$. This proves that $\zeta(s) + \zeta'(s)/\zeta(s)$ has no finite singularities for $\sigma > \frac{1}{2}$. With the known regularity properties of $\zeta(s)$ and the symmetry in zeros from the functional equation (3.2) mentioned above, this proves that no zeros of $\zeta(s)$ can occur in $0 < \sigma < \frac{1}{2}$, nor in $\frac{1}{2} < \sigma$, and so leads to:

<u>Theorem 3.1</u>: <u>The Riemann zeta-function, defined for $\sigma > 1$ as in</u> (3.1) <u>has all its non-trivial zeros on the vertical line $\sigma = \frac{1}{2}$.</u>

The corresponding theorem for the Dirichlet L-functions is proved in analogous fashion. The consequences for analytic number-theory are too well known to be enumerated here.[5]

Possible convergences of (1) on or to the left of $\sigma = \frac{1}{2}$ would not affect RH because singularities of $\xi'(s)/\xi(s)$ would occur in any case on the line $\sigma = \frac{1}{2}$ from the second series on the right in (3.3).

References:

1). Math. Annalen : 99.1928.309-319.

2). As reported by J.V. Uspensky : Introduction to Mathematical Probability (New York, 1937), p.206, ex.17.

3). K.Knopp : Theory and Application of Infinite Series (trans.London 1928), 313-5.

4). E.C.Titchmarsh : Theory of Functions (Oxford,1932); p.153; proofs of function-theoretic results needed for this note will be found in this book.

5). See for example K.Prachar : Primzahlverteilung (Berlin,1957).