Reprinted from "Journal of the Indian Society of Agricultural Statistics," Vol. XI, Nos. 1-2, 1959

AN APPLICATION OF STOCHASTIC CONVERGENCE

BY D. D. KOSAMBI

(*Poona*)

THE basic result of this paper, from which the conclusions of section **3** follow, is the proof that the series

$$\sum_{n \leq x} \frac{1}{n^\sigma} - \sum_{p \leq x} \frac{\log p}{p^\sigma}, \quad x \to \infty, \tag{1}$$

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 $\frac{1}{2} < \sigma < 1$. The classical tools for dealing with such convergence problems[1] are inadequate.

The special device employed in this note resembles to a certain extent the use of Lebesgue integration when the integrand oscillates so much that the evaluation of a Riemann integral does not seem feasible. Only, in place of Lebesgue outer measure, we use probability measure. This is more convenient because of the powerful results in probability theory now available.

The following method is adopted for the proof. Terms of the series (1) are grouped together for n and p in irregular, non-overlapping, consecutive intervals $\boldsymbol{d}_\nu$ of length d_ν, with a uniquely defined, not necessarily integral, real number x_ν in $\boldsymbol{d}_\nu$. The original series (1) is then replaced by

$$\sum_\nu \left\{ \frac{d_\nu}{\log x_\nu} - \pi(\boldsymbol{d}_\nu) \right\} \cdot \frac{\log x_\nu}{x_\nu^\sigma}, \quad \sigma > \frac{1}{2}, \tag{2}$$

where $\pi(\boldsymbol{d}_\nu)$ is the number of primes in $\boldsymbol{d}_\nu$. Differences between the partial sums of (1) and (2) are clearly expressible as the sum of terms

$$\left(\frac{d_\nu}{x_\nu^\sigma} - \sum \frac{1}{n^\sigma} \right) - \left(\pi(\boldsymbol{d}_\nu) \frac{\log x_\nu}{x_\nu^\sigma} - \sum \frac{\log p}{p^\sigma} \right); \quad n, p \subset (\boldsymbol{d}_\nu). \tag{3}$$

Therefore, they may be dissected into components due to the grouping; to the substitution of x_ν for n and p; and those from the partial sums of (1) and (2) not terminating at the same term.

If a set of covering 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, $[d]$ indicates the number of integers in the interval d. The prime number theorem is taken for granted in the form: $\pi(0, x) \sim li(x) \sim x/\log x$. 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 due to A. Kolmogoroff[2] is fundamental:

LEMMA *K.—The stochastic series* $\Sigma \mathrm{u}_n$ *of independent variates* $\{\mathrm{u}_n\}$ *converges with* $\mathrm{P} = 1$ *if there exists another set of independent variates* v_n *such that the series* $\Sigma \mathrm{P}(\mathrm{u}_n \neq \mathrm{v}_n)$, $\Sigma \mathrm{E}(\mathrm{v}_n)$, *and* $\Sigma \mathrm{V}(\mathrm{v}_n)$ *all converge; otherwise the convergence-probability of* $\Sigma \mathrm{u}_n$ *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_ν, $\nu = 1, 2, \cdots$ so that (2) and the series and sequences of its differences with (1) all converge, with a positive compound total probability. Therefore, *at least one infinite sequence of covering intervals* $\{\mathbf{d}_\nu\}$ *must exist giving simultaneous convergence of all these, and* hence the series (1) converges. The existence theorem need not actually construct a specific set $\{d_\nu\}$, 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 the prime numbers been treated as a stochastic sequence because of their irregularity.

In series (2) and some of the associated series and sequences, the occurrence of x_ν makes the terms dependent in probability. This is circumvented by setting up comparison series where the terms are independent, and to which lemma K applies. Similarly, $\pi(d_\nu)$ enters into some of the auxiliary series; there, the probabilities required may be assessed by a change of measure in sample-space. Thereby, the erratic behaviour of the primes in the natural sequence of positive integers, which spoils other proofs, is turned into an asset.

The use of probability methods is easily motivated. If the primes were regularly spaced, $\log p$ apart, (1) would converge for $\sigma > 0$. On the other hand, suppose that $\pi(n) = li(n)$ exactly, whenever $n = k^a$,

k an integer and $\alpha \geq 3$. Further, let these 'pseudo-primes' cluster together at the left-hand end of $k^{\alpha} \leq n < (k+1)^{\alpha}$, leaving the rest of the interval void of primes. Then the corresponding series (1) clearly diverges for some $\sigma > \frac{1}{2}$. For the actual primes, the gaps would be much too large even when $\alpha = 2$ (which does not give divergence); for, it is known[4] that for almost all x and h of order $x^{\frac{1}{4}}$, $\pi(x+h) - \pi(x) \sim h/\log h$. Thus, known facts about primes suffice to exclude regular arrangements that would make (1) diverge. The question is really settled by showing that the prime numbers in suitably defined intervals behave like an *unbiased random sample* from a non-singular probability distribution (or like a von Mises *Kollektiv*). That is, the relative frequencies of intervals containing 0, 1, 2, $\cdots$ primes each tend to definite limits as the real line is progressively covered. This is shown in lemma 1·2, which should be the most useful result of this note.

The problem is *to discover an underlying stochastic population* from which an irregular infinite sequence, specified by a procedure, not by formula, might be drawn as an *unbiased* random sample. The answer can be obtained only in the sense of unit probability. For the sequence of primes considered directly, the question of bias would still remain, *i.e.*, whether they do not form part of the exceptional set of zero probability measure. We consider instead an infinite set of complete coverings defined by choice of the initial point, the situation repeating itself when the initial point moves through a single covering interval. Unit probability would mean 'for almost all initial points'. This validates applications of the basic Poisson distribution, which is even more important and useful than the Riemann Hypothesis.

1. This section deals with the mechanism of choice for $\boldsymbol{d}$. Text-book[3] results in probability theory are taken for granted. The real half-line $2 \leq x \leq \infty$, on which the integers and primes are marked off, is transformed into $2 \leq y \leq \infty$ by $y = li(x) + c = \int dx/\log x$. Then, to an interval $(a, a+\Delta)$ on the y-line corresponds a unique interval $\boldsymbol{d}$ on the x-line and conversely, with $\Delta = d/\log x$; where x is chosen as that number (not necessarily an integer) lying in $\boldsymbol{d}$, which makes this relationship hold. The mean-value theorem for the integral of a monotonic function shows the existence of such an x, which lies properly within $\boldsymbol{d}$. 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.

Each length Δ_{ν}, $\nu = 1, 2, \cdots$ is taken to have the identical distribution, namely, the uniform distribution over (0, 2). Being open on the right, marking off consecutive 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 the uniform distribution; the position of the interval of length Δ_ν is uniquely determined by the particular sample. Hence the d-intervals that correspond by the inverse $li\,(x)$ transformation give a stochastic covering of the x-line, one complete covering for each such infinite 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:

$$E\,(\Delta) = 1; \quad V\,(\Delta) = \frac{1}{3}; \quad E\,(\overline{\Delta - 1}^{\,2k}) = \frac{1}{(2k+1)};$$

$$E\,(\overline{\Delta - 1}^{\,2k+1}) = 0. \tag{1.1}$$

The variate y_ν is defined as the sum of the first ν independent, consecutive, non-overlapping Δ-intervals: $y_\nu = \Delta_1 + \Delta_2 + \cdots + \Delta_\nu$; it has the range $(0, 2\nu)$, mean ν, and variance $\nu/3$. Its probability curve is convex upwards, with a single maximum. According to the central limit theorem, the probability distribution of y_ν is approximated efficiently by a normal (Gaussian) distribution with mean ν and standard deviation $\sqrt{\nu/3}$. It follows that the maximum height of the y_ν probability curve is rapidly asymptotic to $1/\sqrt{2\pi\nu/3} = a/\sqrt{\nu}$ and the distribution may be taken as approximately uniform over steps of order less than $\sqrt{\nu}$ in width. The estimates of S. Bernstein[3] may be applied to (1.1) to give:

LEMMA 1.1.—*The probability is less than exp.* $(-\,t^2)$ *for each of the inequalities to hold (separately) for all large* ν:

$$y_\nu > \nu + t \cdot \sqrt{\frac{2\nu}{3}} \quad \textit{and} \quad y_\nu < \nu - t \cdot \sqrt{\frac{2\nu}{3}}. \tag{1.2}$$

Two useful corollaries follow. Taking $t = \sqrt{(3/2)\log \nu}$, *P is less than* $\nu^{-3/2}$ *for each of the inequalities*

$$x_\nu > 2\nu \log \nu\,; \quad x_\nu < \frac{\nu}{2}, \tag{1.3}$$

for all large x *and* ν. *Secondly, the ratio* y_ν/ν *converges in probability to unity.*

LEMMA 1.2.—$\pi\,(\mathrm{d})$ *has a proper frequency distribution over almost all complete coverings, the expectation being unity, and the variance* *finite. If consecutive intervals be grouped together* k *at a time, then the mean and the variance of primes covered are each multiplied by* k.

Proof.—The prime number theorem and lemma 1.1 show that the limit as $r \to \infty$ of $\{\pi(d_k) + \pi(d_{k+1}) + \cdots + \pi(d_{k+r})\}/r$ is unity, with unit probability, for every k. But this only gives the general expectation for almost all complete coverings. The limiting distribution may be singular if the primes occurred in maximal clumps separated by sufficiently many voids to restore the average. In the limit, the frequency of intervals with no primes could then be unity, all others zero—and yet no finite limiting variance need exist. Known results on gaps between successive primes (Prachar,[4] p. 154 *ff.*) make this singular case very unlikely, but we need only appeal to the principle of the sieve method. If all multiples of 2, 3, 5 $\cdots$ are successively struck out, the smallest integer left at each stage is itself the prime to be used in the next deletion; and every prime is reached in this way. The survivors are thus asymptotic to $n(1 - 1/2)(1 - 1/3)\cdots(1 - 1/p)$ ....where the product must be suitably terminated. This says precisely that the (suitably bounded) primes act, each with its own probability $1/p$, *independently* (or the probabilities for survival would not be multiplied as above) of each other in the deletion. There is no linear (or even algebraic) relationship between the primes; and any two or more primes have the highest common factor one while $p_k \sim k.\log k$. The theorem of de la Vallée Poussin says that for any arithmetic progression $ar + b$, $r = 1, 2, \cdots$ the primes are asymptotically equally divided between the $\phi(a)$ possible different categories, no matter what a is chosen.

Therefore, the number of primes 'striking' an integer and the number of integers escaping the sieve ought each to have some sort of asymptotic frequency distribution. Of these, the first is given by Landau's theorem, that the relative frequency of integers $< n$ having $k + 1$ prime factors is asymptotic to $e^{-t}t^k/k!$ for $k = 0, 1, 2\cdots$, with $t = \log\log n$. This is a Poisson distribution, and the value of the parameter t would follow from the prime number theorem, if the distribution were granted.

For the prime survivors, we first take an interval $\boldsymbol{h}$ of y-image μ (fixed), hence of x-length approximately $\mu.\log n$. This is allowed to cover, with a uniform probability, the total range whose image of length $N(n)$ contains the integer n. Then all deletions from $\boldsymbol{h}$ may be considered as due to primes not exceeding $\sqrt{n + N}$. Of these, the primes smaller than h will cause compulsory deletions, but those between $\mu.\log n$ and $\sqrt{n+N}$ will act as a matter of chance. The survivors of the compulsory deletions are $\approx e^{-\gamma}h/\log h$, (where γ is Euler's constant) which increases beyond any limit. The mean being μ, the probability of the survivors of the first deletion in the interval containing a prime will be $\approx \mu\, e^{\gamma} \log h/h$, which tends to zero. The introduction of probability methods is needed because not enough is known about the location of primes for direct calculation of their frequency distribution. This does not distort the actual distribution, particularly as regular arrangements have been disposed of in our preamble. The μ-intervals still belong to a covering, hence do not overlap. It suffices, therefore, to choose any interval at random after giving the initial point of the first μ-interval in the range a uniform distribution over one interval length. The number of primes neglected cannot exceed the maximum covered by a μ-interval of the N-range; which affects neither the distribution nor the convergence of (1).

Statements about the number of 'survivors' per interval have to be understood in the sense of unit probability. The arrangement of integers not divisible by $2.\ 3.\ 5.\ \cdots, p_r$ is repeated modulo $N_r = 2.\ 3.\ 5.\ \cdots, p_r$. The theorem of Mertens gives the proportion of numbers prime to N_r as $\sim e^{-\gamma}/\log p_r$. Only the very small primes with $N_r \leqslant h = \mu \log n$ can have a cyclic effect over an interval length. The less regular effect of the remaining small primes $\leqslant h$ is most economically described as an independent survival probability $\sim \log p_r/\log h$, because the initial point and length of the range are each of order $exp\,(cp_r)$. By classical probability theorems, the chance of an interval having less than $ah/\log h$ survivors (with suitable $a > 0$) tends to zero. However, every one of the consecutive integers $kN_r + 2, kN_r + 3, \ldots, kN_r + p_{r+1} - 1$ has a factor in common with N_r. Inasmuch as N_r is of order n^{μ}, there could be (for small μ) intervals in a range devoid of survivors. These, or intervals with less than any fixed number of survivors, may be ignored in the limiting process as zero-probability phenomena.

We have now to consider whether the chances of an integer being a prime or composite are affected by the knowledge that some other integer in the interval is actually a prime, or composite. Should r be a prime > 3, then $r + 1$ and $r - 1$ are necessarily composite. Such obligatory dependence is removed by striking out the multiples of 2, $3, \cdots p < \mu \log n$. Suppose that among the 'first survivors' one is known to be composite. Then its prime factors cannot, by construction, divide any other in the same μ-interval. The chance P' for primality among the rest may at worst have to be P (the original probability) multiplied by $1/(1 - 1/p)$ for every such prime factor. By Landau's theorem above, the average number of prime factors is of order $\log_2 n$; the maximum number of such factors can obviously not exceed $\log n/\log_2 n$. Thus P would at most have to be multiplied by $e^{\phi(n)}$, where $\log_2 n < 1/\phi(n) < \log n$. The supply of primes which cause deletion, being of order $\sqrt{n}/\log n$, is not materially depleted, so that the argument may be repeated for further numbers found composite. On the other hand, if the known integer be a prime, the probability for the rest is not thereby affected in the same interval, for deletion is caused only by primes $< \sqrt{n}$, approximately. Thus any modification of P is an infinitesimal of higher order, which justifies passage to the limit on the basis of independence in probability among the 'first survivors' within the interval. This is also supported by known sieve theorems (Prachar,[4] Chap. II). Parallel arguments hold *a fortiori* for independence between intervals. As has been shown above, the number of these survivors tends to infinity with n, P tending reciprocally to zero, while the expectation is μ. It follows[7] that the limiting distribution is Poissonian provided there are an unboundedly increasing number of disjoint intervals in the range and $E\{\pi(h)\} \to \mu$ over the separate ranges as $n \to \infty$. These conditions are met by taking $N = n^{\frac{3}{4}}$, though smaller exponents will do as well.[4] The distribution therefore approximates rapidly to the frequencies $e^{-\mu}(1, \mu, \mu^2/2 \cdots)$ for $\pi(h) = 0, 1, 2 \cdots$. If, instead of a fixed length μ, we allowed a uniform distribution over $(0, \mu)$ for the length, simple integration would yield the asymptotic frequency for k primes as $(1 - e^{-\mu} s_k)/\mu$, where s_k is the sum of the first $k+1$ terms in the Maclaurin expansion of e^{μ}. The mean is now $\mu/2$, and the variance becomes $(\mu/2 + \mu^2/12)$, the second term being the "Sheppard's correction for grouping" familiar to statisticians.

For any finite number of consecutive N-ranges and fixed μ, the distribution is the weighted average of the component distributions over each range, with the number of μ-intervals in the range as weight. The whole line being thus progressively covered, this amounts to summability by a *regular* Toeplitz matrix, of the sequence of range-distributions. Therefore, *the distribution over the whole line is the same as the asymptotic distribution over the* N-*range, namely, Poissonian with mean and variance* μ. The other distribution derived naturally holds over the complete real line also. The Poisson distribution being valid for any μ, grouping consecutive intervals together k at a time (every interval belonging to one and only one such grouping) again gives a Poisson distribution with parameter $k\mu$. With uniform distribution of interval length over $(0, \mu)$ the mean and variance of primes will be $\frac{1}{2}k\mu$ and $k(\mu/2 + \mu^2/12)$ respectively; in our special case, k and $4k/3$. Thus, *the* $\pi(\boldsymbol{d})$ *in consecutive intervals (of a complete covering) grouped together add like independent random variables.* Q.E.D.

The distribution itself is less important than the existence of a non-singular distribution. Each individual $\pi(\boldsymbol{d}_\nu)$ has also some frequency distribution for fixed index ν, which obviously tends to the distribution over a complete covering as the index increases beyond limit.

A more number-theoretic proof of this fundamental lemma would run as follows: Brun's sieve theorem extends (*note* 4, *p.* 52, *th.* 4.7) to: *The number of primes* $p \leq N$ *for which* $p + b_1, p + b_2, \cdots, p + b_r$, $0 < b_1 < b_2 < \cdots, b_r$, *are also primes is less than* $cMN/\log^{r+1} N$, *where* $M \leq \Pi(1 - 1/p)^{-r}$, *taken over all primes dividing* $\Pi b_i \Pi (b_i - b_k) > 0$. Let $p, p + b_i$ be restricted to lie within a single covering interval of length $h \leq \mu \log N$. By the theorem of Mertens, $M < a^r \log^r h$, where a is a constant. The b_i can be chosen at will provided there is no *a priori* restriction to prevent $p + b_i$ being a prime; *e.g.*, b_i must be even for $p > 2$. This means precisely that every $p + b_i$ must be a 'first survivor'. If R be the number of choices for any b_i, $R < Bh/\log h$. This follows in the sense of unit limiting probability from the preceding paragraphs, while it is known from purely number-theoretic considerations that no interval of length f can contain more than $Df/\log f$ primes, D constant. The whole set of b's may be specified in $R!/(R - r)!\, r!$ different ways. The number of covering intervals being N/h, the relative frequency of intervals covering $r + 1$ primes cannot exceed the binomial coefficient above, multiplied by $CM/\log^r N$. Therefore, ultimately, *the frequency of intervals having* $r + 1 \geq 2$ *primes cannot exceed* $Ab^r/r!$ *(A, b const.).* This suffices to prove the existence of the second and higher moments, but the vital Poisson distribution would require further refinement of the sieve, or probability arguments.

2. The series (2) is now written as the difference of the two stochastic series, whose convergence is to be considered separately:

$$\Sigma_\nu (\Delta_\nu - 1) \cdot \frac{\log x_\nu}{x_\nu^{\sigma}} - \Sigma_\nu [\pi(\boldsymbol{d}_\nu) - 1] \cdot \frac{\log x_\nu}{x_\nu^{\sigma}}\,; \quad \sigma > \frac{1}{2}. \quad (2.1)$$

THEOREM 2.1. *The series* $\Sigma_\nu (\Delta_\nu - 1) \,.\, \log x_\nu / x_\nu^{\sigma}$ *converges with unit probability for* $\sigma > \frac{1}{2}$ *and zero probability for* $\sigma < \frac{1}{2}$.

Proof: *Step* 1.—There exists a critical value σ_0 of the exponent σ such that the convergence probabilities for the series under consideration are $P = 0$ for $\sigma < \sigma_0$ and $P > 0$ for $\sigma > \sigma_0$. For, if any series with a specific choice of $\boldsymbol{d}$ converge for a given σ, it necessarily converges for all greater values of σ; if it diverge, then divergence follows for all lesser values of the exponent. This is a consequence of standard results in the theory[1] of infinite series, noting that the coefficients outside the brackets are ultimately monotonically decreasing and positive. Thus, if the convergence probability be zero for any exponent, it cannot be positive for any lesser value of σ. This enables a Dedekind section to be defined for the values of σ, between the zero and the non-zero probability ranges.

Step 2.—The convergence exponent is the same when $\log x_\nu / x_\nu{}^\sigma$ is replaced by $\log \nu/\nu^\sigma$ in the coefficients. To prove this, we note that by lemma 1.1, an arbitrarily small $\epsilon > 0$ may be chosen, with two suitable sets of positive constants a, b; $\bar{a}, \bar{b}$ such that $\log x_\nu / x_\nu{}^\sigma$ is bracketed between $a \,.\, \log \nu^b/\nu^{\sigma+\epsilon}$ and $\bar{a} \,.\, \log \nu^{\bar{b}}/\nu^{\sigma-\epsilon}$. Moreover, the probability P_ν for each bracketing is such that ΠP_ν converges. Conversely, a similar bracketing of $\log \nu/\nu^\sigma$ by corresponding terms in x_ν is also obviously possible. [In each case, the *log* terms in the factors may be ignored, as $\log^k z = 0\,(z^\epsilon)$ for every k and every positive ϵ.] It follows that if $\sigma < \sigma_0$ in the x-series, the convergence probability cannot be positive for the ν-series with the same exponent; similarly for $\sigma > \sigma_0$.

Step 3.—Lemma K applies to the series $\Sigma_\nu (\Delta_\nu - 1) \,.\, \log \nu/\nu^\sigma$. The term means are all zero; the variances are $\log^2 \nu/3\nu^{2\sigma}$. The critical exponent for the ν-series, and therefore the x-series also is thus $\sigma = \frac{1}{2}$. Finally, the convergence probability for the x-series with $\sigma > \frac{1}{2}$ is at least ΠP_ν. Inasmuch as $x_\nu \geq 2$, the contribution from any finite number of initial terms remains finite, regardless of probability considerations, and the terms may be omitted without affecting convergence. However, the omission of the corresponding terms in ΠP_ν brings the probability arbitrarily near to unity. Hence, the probability of convergence for $\sigma > \frac{1}{2}$ must be unity even for the x-series. This completes the proof, though it suffices for our ultimate purpose that $P > 0$ for $\sigma > \frac{1}{2}$.

THEOREM 2.2. *The series* $\Sigma_\nu [\pi(\mathbf{d}_\nu) - 1] \,.\, \log x_\nu / x_\nu{}^\sigma$ *has likewise* $\mathrm{P} > 0$ *for convergence when* $\sigma > \frac{1}{2}$.

Proof.—The existence of a critical exponent which coincides with that of the comparison series $\Sigma_\nu [\pi(d_\nu) - 1] \,.\, \log \nu/\nu^\sigma$ is proved as in the preceding theorem, so that we may deal only with the latter series.

Lemma 1.2 shows that for almost every complete covering $\pi(\boldsymbol{d}) - 1$ has a zero mean (*over the covering*, not for fixed index ν) and finite variance. Therefore, with probability P_0 arbitrarily close to unity, $V\{\pi(\boldsymbol{d})\} < A^2$ over complete coverings for suitably large A. The sum of any m terms $\pi(\boldsymbol{d}) - 1$ taken at random from the same covering would be less than $A.t.\sqrt{m}$ in absolute value, with $P > 1 - 1/t^2$. Take non-overlapping consecutive blocks of $m = 2^k$ consecutive terms of the covering, with $k = 2, 3, \cdots$ and the understanding that an arbitrary number of initial terms of the series may eventually be omitted at need. The first subscript ν in each block will be equal to the total number of terms in that block. By Abel's lemma[1] and taking $t = k$, the sum of the terms in the comparison series corresponding to the k-th block will be less absolutely than $Ak^2 \log 2/2^{k \cdot \epsilon}$ (where $\epsilon = \sigma - \frac{1}{2} > 0$) with probability $P_k > 1 - 1/k^2$. Hence the series has a convergence probability not less than $P_0 \Pi P_k > 0$.

Choice of consecutive instead of completely random intervals does not vitiate the result. For, the existence of a distribution for $\pi(\boldsymbol{d})$ was proved as for consecutive intervals; and the selection is uninfluenced at any stage by the actual prime content of any intervals or blocks. In fact, it is known[4] that if m consecutive intervals together cover a stretch of magnitude y^c for any $c > 38/61$, then the (stochastic) block sum under consideration is $0(2m)$ as $y \to \infty$, without any probability condition or exceptional set of integers. This is stronger than what is demanded or yielded by probability considerations. That is, the values assumed by sums of $\pi(\boldsymbol{d}) - 1$ for sufficiently great block lengths cannot be more extreme for consecutive covering intervals than with random choice. Lemma 1.1 says, however, that every block length will be of order arbitrarily close to y with a compound probability given by an infinite product that converges to some $P^* > 0$. Thus the inequalities can be strengthened, with a convergence probability at the worst multiplied by another factor P^*. The critical exponent of convergence for the series remains $\sigma_0 = \frac{1}{2}$.

Though justified by lemma 1.2, the multiplication of probabilities P_k is unnecessary. No matter what the joint probabilities, the chance for *all* the grouped sums lying within the absolute limits given above for each is not less than $1 - \Sigma(1 - P_k)$, or than $1 - \Sigma 1/k^2$, which can be brought arbitrarily close to unity by rejection of enough initial terms of the series. The grouping of intervals need not be in geometric progression. It would suffice to combine $(k + 1)^a - k^a$

successive intervals at a time with $k = 1, 2, \cdots$ and $\alpha > 3/(2\sigma - 1)$. The process need begin only from some $\nu = \nu_0$. Thus far, only the simple Chebyshov inequality is used, which requires nothing beyond the existence of a distribution with finite second moment. Use of the actual distributions found in lemma 1.2, with or without the normal approximation given by the Central Limit Theorem, permits still freer groupings. Q.E.D.

An alternative proof of theorem 2.2 would run as follows. The correspondence $\boldsymbol{d} \to \pi(\boldsymbol{d})$ maps the space of all permissible complete coverings into the points of an integral infinite-dimensional lattice with co-ordinates $x_i = \pi(\boldsymbol{d}_i)$. We take the lattice as right-angled, and redefine the measure by giving equal weight to every point actually realized. Suppress a suitably large but finite number of initial dimensions altogether. Then, if necessary, trim off just enough peripheral points to make the centre of gravity (with equal weights) the unit point $(1, 1, 1, \cdots)$. This can always be done with an arbitrarily small measure of deletion because almost every realizable point of the lattice has $(x_1 + x_2 + \cdots + x_n)/n \to 1$ and a limiting distribution (in the old measure) is approached by all x_i with large i which is also the distribution over the successive x-co-ordinate values of almost every point. What is left may be further restricted to almost all lattice points of an infinite-dimensional hypercube (with unequal indefinitely increasing sides), of the same centre of gravity, and with a rectangular section in any finite number of dimensions. For, if $x_i = a$ is realizable, then so is every value $0 \leq x_i \leq a$, by contraction of interval length.

The new measure is defined over this lattice hypercube as the proportion of points lying in any included region to be measured. The total measure of the hypercube is unity, with an induced measure over every subspace which is defined as the relative number of points in the lattice hypercube lying in the cylinder with a region of the subspace as base and sides extended over the entire ortho-complement. Then the measure over the product-space is the product of the component measures. This change of probability-measure amounts to the integration of a positive weight function. It suffices for our purpose that no set of measure unity in the new lattice measure is of measure zero in the original measure. In the lattice measure, the variates $\pi(\boldsymbol{d}_\nu)$ become independent in probability; each has a unit mean because of the centre of gravity chosen, and the variance can never be greater than $2.\log^2 \nu$, whatever the actual distribution. Therefore, lemma K becomes immediately applicable, and the comparison series converges with uniform lattice measure one for every $\sigma > \frac{1}{2}$; hence the original series with $P > 0$ in the previous measure.

THEOREM 2.3. *The difference between the series* (2) *and* (1) *may be resolved into the following series and sequences, each of which converges with unit probability for* $\sigma > 0$:

$$\sum_{\nu} \frac{(d_\nu - [d_\nu])}{x_\nu^{\sigma}} ;$$

$$\sum_{\nu} \frac{\pi(d_\nu).d_\nu(\sigma.\log x_\nu - 1) - \sigma.d_\nu(d_\nu + 1)}{x_\nu^{1+\sigma}} ;$$

$$\frac{(d_\nu + 1)}{x_\nu^{\sigma}} ; \qquad \pi(d_\nu) \cdot \frac{\log x_\nu}{x_\nu^{\sigma}} . \tag{2.3}$$

Proof.—The first of these is due to there being $[d]$ and not d integers in $\boldsymbol{d}$. Now not only $d - [d]$ but the sum of any number of such differences for consecutive non-overlapping intervals $\boldsymbol{d}_\nu$ ranges by definition between -1 and $+1$ without reaching either extreme. By Abel's lemma, the series will converge provided $x_\nu \to \infty$ monotonically, for which the probability is unity. Terms of the second series may be compared with $\log^3 \nu / \nu^{1+\sigma}$; and so it also converges with $P = 1$ for $\sigma > 0$. The two sequences are due to the partial sums of (1) and (2) not necessarily terminating at the same place: both obviously converge with $P = 1$ for $\sigma > 0$. Q.E.D.

The three auxiliary theorems lead immediately to:

THEOREM 2.4. *The series* (1) *converges for* $\sigma > \frac{1}{2}$.

Proof.—The series (1) converges if and only if (2) and (2.3) converge for at least one choice of consecutive non-overlapping intervals $\boldsymbol{d}$. If no such choice exists, the joint probability for the simultaneous convergence of all the stochastic series and sequences in (2) and (2.3) would have to be zero. But the joint probability is positive (in fact arbitrarily close to unity). Q.E.D.

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

$$\zeta(s) = \sum_{1}^{\infty} \frac{1}{n^s} = \Pi \frac{1}{(1 - p^{-s})} . \tag{3.1}$$

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

The zeta-function obeys the functional equation[5]:

$$\zeta(1-s) = 2^{1-s}\pi^{-s}\cos\left(\frac{\pi s}{2}\right)\Gamma(s)\,\zeta(s). \tag{3.2}$$

The Riemann hypothesis (*RH*) is the conjecture that all zeros of $\zeta(s)$ not $s = -2, -4, \cdots$ lie on the vertical line $\sigma = \frac{1}{2}$. It is easily seen, directly from the convergence of the infinite product, that no zero can occur in $\sigma > 1$. It is also known from a theorem of G. H. Hardy that an infinity of zeros lie on the line $\sigma = \frac{1}{2}$. Using the functional equation, it would suffice to prove *RH* if it could be shown that no zero lies in the critical half-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* $F(z) + F'(z)/F(z)$. Only the simple poles $1/(z-a)$ cancel out, but zeros of $F(z)$ now appear as first degree poles because of the second term, the logarithmic derivative. For $F(s) = \zeta(s)$, the fact that $\zeta(s)$ has no finite singulariy 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 $\log(1-x) = -x - x^2/2 - x^3/3 - \cdots$:

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_p \frac{\log p}{p^s} + \sum_p \frac{\log p}{2p^{2s}} + \sum_p \frac{\log p}{3p^{3s}} + \cdots. \tag{3.3}$$

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 $2\Sigma \log n/n^{2\sigma} = -2\zeta'(2\sigma)$. Therefore, the discussion by means of $\zeta(s) + \zeta'(s)/\zeta(s)$ reduces to showing that the Dirichlet series

$$\sum \frac{1}{n^s} - \sum \frac{\log p}{p^s}, \quad p, n \leqslant x \to \infty, \quad s = \sigma + it, \tag{3.4}$$

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 $\sigma > \frac{1}{2}$. Hence, by the known property of such Dirichlet series, (3.4) converges uniformly in any half-plane to the right of $\sigma = \frac{1}{2}$. This proves that $\zeta(s) + \zeta'(s)/\zeta(s)$ has no finite singularities for $\sigma > \frac{1}{2}$. Therefore, no zeros of $\zeta(s)$ can occur in $\frac{1}{2} < \sigma$, proving *RH*:

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

The corresponding theorem for the Dirichlet L-functions is proved in analogous fashion. The consequences are well known.[6]

Possible convergence of (1) on or to the left of $\sigma = \frac{1}{2}$ would not affect *RH* because singularities of $\zeta'(s)/\zeta(s)$ occur in any case on the line $\sigma = \frac{1}{2}$ from the second series on the right in (3.3). Moreover, the function $Q(s) = -\Sigma' \log p/p^s$ is $\Sigma \mu(n) \zeta'(ns)/\zeta(ns)$ in $\sigma > 0$, so has poles on $\sigma = \frac{1}{2}$. Hence (1), the Dirichlet series for $\zeta(s) + Q(s)$, cannot converge beyond the critical line.

The probability approach allows some conclusions to be drawn quickly without the intermediacy of *RH*. For example, the Poisson distribution for primes covered by unit-image intervals, and the famous law of the iterated logarithm allow a probability estimate of $|\pi(x) - li(x)|$. This, under the assumption of independence for primes in the given intervals, would exceed with probability arbitrarily close to unity, the magnitude $(1 - \delta)\sqrt{2y.\log\log y}$. The probability would be arbitrarily close to zero if the $-\delta$ be replaced by $+\delta$. With $y \sim x/\log x$, it is seen that the original Littlewood result is not quite the best possible.

The zeros of $li(x) - \pi(x)$ appear as recurrence times (on the y-scale) for the equilibrium of a Possion variate. The distance between consecutive primes amounts to the 'waiting time' on the y-scale, and has a distribution given by $dP = e^{-\mu} d\mu$. It follows that *for any* $\phi(\mathrm{n}) = o(\log \mathrm{n})$ *and infinitely many primes* p, *the separation from the next prime will exceed* $\phi(\mathrm{p}) \log \mathrm{p}$. Systematic use of the Poisson distribution would eliminate theorem 2.1 altogether, but would not bring out the basic fact that theorem 2.2 is independent of any reasonable choice of covering intervals. Finally, *RH* may be generalized to Dirichlet series whose exponents (our $\log n$) form a complete Abelian semi-group under addition with a basis set of generators, our $\log p$. But no generalization of *RH* exists if the product corresponding to $\Pi(1 - 1/p^s)$ converges in the half-plane $\sigma > 0$. This covers the case where the generator basis is finite, and should explain the negative Bourbaki-Weil result for Abelian fields.

REFERENCES

1. Knopp, K. .. *Theory and Application of Infinite Series* (Trans. London, 1928), particularly pp. 313–15.

2. Kolmogorov, A. .. *Math. Annalen*, 1928, **99**, 309–19.

3. Uspensky, J. V. .. *Introduction to Mathematical Probability*, New York, 1937, contains the elementary probability theory utilized here. For the Bernstein estimates, *see* pp. 204–06. Also, reference 7 below.

4. Prachar, K. .. *Primzahlverteilung*, Berlin, 1957, pp. 8, 157–64, etc. The position of *RH* is made clear in various contexts throughout the book. For the Selberg asymptotic result, pp. 323–24; without an exceptional set of integers of measure zero, the best known exponent is 38/61, reducing that given by A. E. Ingham in *Quart. J. Math.*, Oxford, 1937, pp. 255–66.

5. Titchmarsh, E. C. .. *Theory of Functions*, Oxford, 1932, p. 153; Chapter IX, pp. 113–16, and p. 260 for other function-theoretic results used here. See also the next reference.

6. ———— .. *Theory of the Riemann Zeta-function*, Oxford, 1951, Chapters, XIII–XIV; also, Prachar as in note 4 above.

7. Feller, W. .. *An Introduction to Probability Theory and Its Applications*, Vol. 1, New York, 1950, sections 6.4, 6.5, 11.8 and 17.2; further developments have been promised for the second volume. See also H. Richter: *Wahrscheinlichkeitstheorie* (Berlin, 1956), 336–40. For extensions of the Chebyshov inequality, *Ibid.*, 256, though not used here.