NUMBER FIFTY-TWO

The retiring editorial staff nursed from cradle through publication ten MONTHLYS per year for the last five years, also two Slaught papers. We hope most have been satisfactory, some excellent. Our debt to our authors and referees is great, also to our readers for their steady support and encouragement.

I personally am grateful to all of my associate and collaborating editors. They worked hard and competently to a high professional standard. They join me in wishing our successors well.

Harley Flanders

PRIME NUMBERS AND BROWNIAN MOTION

PATRICK BILLINGSLEY, The University of Chicago

Because it factors into a product of prime numbers, each integer contains within it a kind of Brownian motion path, and the mathematics of Brownian motion can be used to derive theorems about the factorization. Despite the persistent notion that a result stated in probability language is rather less true than it might otherwise be, I shall state these theorems in probability language and even give them probabilistic proofs. As a matter of fact, there will be little in the way of real proofs, since for the most part I shall only illustrate general results by examples and special cases. For this there is the authority of William Feller, who used to tell us, his students, that the best in mathematics, as in art, letters, and all else — that the best consists of the general embodied in the concrete. Although at first I thought that was simply an antimilitary sentiment, I did eventually understand it as the intellectual-esthetic principle he intended and have tried ever since to keep it at the front of my mind.

The paper has three sections. In Section 1, the mathematical model for a particle in Brownian motion is defined and some of its properties described. Section 2, which provides the link between Brownian motion and primes, concerns random walk: one successively tosses a coin and successively moves along a scale, one unit in the positive

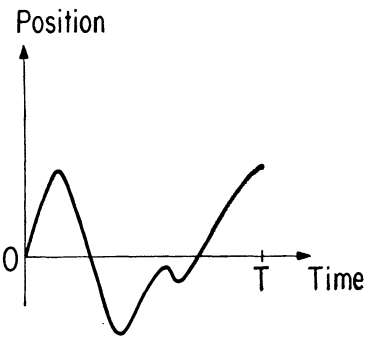
Patrick Billingsley received his Princeton Ph.D under William Feller. Except for a period of Navy service, he has been at the University of Chicago, where he is presently Professor of Mathematics and Statistics. He was a Fulbright lecturer for a year at the University of Copenhagen, and held a Guggenheim fellowship at Cambridge University. Professor Billingsley's main research interest is probability theory, and he is a fellow of the Institute of Mathematical Statistics. His publications include the books: *Statistical Inference for Markov Processes*, 1961, *Ergodic Theory and Information*, 1965, *Convergence of Probability Measures*, 1968, *Elements of Statistical Inference* (with D. L. Huntsberger), 1973. *Editor.*

or negative direction, according as the coin falls heads or tails. Here it is shown how a distant random walk looks approximately like a Brownian motion and how the Brownian motion model therefore leads to limit theorems associated with random walk. Section 3 discusses the random walk which a randomly chosen integer generates through its prime factorization: one successively examines the primes, 2, 3, 5, ⋯, and successively moves along a scale, one unit in the positive or negative direction, according as the prime appears in the factorization or not. It turns out that, because of the arithmetic fact that distinct primes individually divide an integer if and only if their product does, this factorization random walk has many of the properties of the ordinary coin-tossing random walk; in particular, it too can be approximated by Brownian motion, and it is shown how this leads to limit theorems associated with factorization into primes.

In addition to the elements of real analysis, the paper makes use of statistical concepts such as mean, variance, independence, and Gaussian distribution.

1. Brownian motion. Imagine suspended in a fluid a particle bombarded by molecules in thermal motion. The particle will perform that irregular and seemingly random movement first described by the biologist Robert Brown in 1828. Since we shall be concerned with just one component of this motion, imagine it projected on a vertical axis: At each instant t of time we note the height $x(t)$ of the particle above a fixed horizontal plane. Over T units of time, the motion of the particle, which we take to start at 0, is described by the positions $x(t)$ for $0 \leqq t \leqq T$—that is, by a continuous real function x on $[0, T]$ with $x(0) = 0$. This leads us to consider the collection $C_0[0, T]$ of such functions x.

FIGURE 1

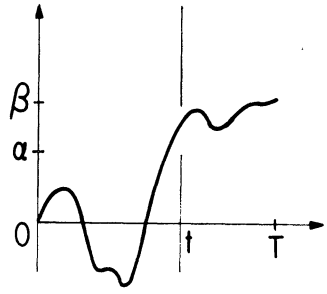


[1] For technical reasons, we make $C_0[0, T]$ into a metric space by taking the distance between two of its elements to be the maximum vertical distance between their graphs. This topology, the uniform topology, is of little direct concern here; it is brought in mostly as evidence that the discussion to follow does have a rigorous basis.

The random motion of the particle is described by an assignment of probabilities

$P_T(A)$ to subsets A of $C_0[0, T]$; $P_T(A)$ represents the chance that the path traced out by the particle lies in A, or is described by a function x that lies in A. Probabilities represent long-run relative frequencies. If the total on a pair of dice is observed, the possible outcomes are $2, 3, \cdots, 12$. If many pairs of balanced dice are rolled independently, the proportion among them producing the outcome 7 will be about $1/6$. If a particle in Brownian motion is observed for T units of time, the possible outcomes are the various elements of $C_0[0, T]$. If many independently moving particles are observed, the proportion among them producing paths that lie in A will be about $P_T(A)$. Although the interpretation of probability involves such multiple observations, in the mathematical theory we speak of a single roll of the dice, the probability the roll produces a 7 being $1/6$; in the same way, we speak of a single particle, the probability it traces out a path that lies in A being $P_T(A)$.

FIGURE 2



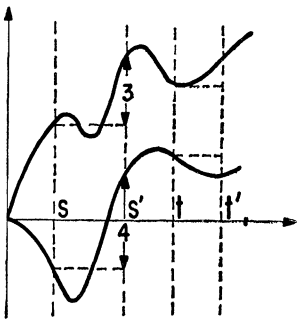
The set $[x: \alpha \leqq x(t) \leqq \beta]$, consisting of the paths that go through the gate in Figure 2, represents the event that at time t the particle will lie between α and β; it is assigned probability

$$P_T[x: \alpha \leqq x(t) \leqq \beta] = \frac{1}{\sqrt{2\pi t}} \int_\alpha^\beta e^{-u^2/2t}\, dt. \tag{1}$$

Thus the distribution of the position at time t follows the Gaussian curve with mean 0 and variance t. That the mean is 0 reflects the fact that the particle is as likely to go up as to go down; there is no drift. The variance t grows linearly; this indicates that the particle tends to wander away from its starting point and, having done so, suffers no force tending to restore it to that starting point. The equation (1) can be extended: the increment over $[s, t]$ has a Gaussian distribution with mean 0 and variance $t - s$.

The other important property of Brownian motion is this: Suppose $s < s' < t < t'$, and consider for example the event $A = [x: x(s') - x(s) \geqq 3]$ that the particle undergoes an upward displacement of at least 3 units during the time interval $[s, s']$, together with the event $B = [x: x(t') - x(t) < 0]$ that the particle undergoes a

FIGURE 3



downward displacement during the time interval $[t, t']$. The top path in Figure 3 lies in A but not in B, and the bottom path lies both in A and in B. The probabilities of A and B and of their intersection $A \cap B$ are related by

$$P_T(A \cap B) = P_T(A)P_T(B). \tag{2}$$

Thus A and B satisfy the definition of independence; that is, that the displacement the particle undergoes during $[s, s']$ in no way influences the displacement it undergoes during $[t, t']$. This implies a kind of lack of memory. Although the future behavior of the particle depends on its present position, it does not depend on how the particle got there. Equation (2) has a more general form showing that the increments over any number of disjoint intervals are statistically independent of one another.

The equations (1) and (2), together with generalized versions of them, determine all the probabilities $P_T(A)$. (This ignores a technical point: $P_T(A)$ cannot be defined for every subset A of $C_0[0, T]$, but it can for every Borel set A — that is, for every A in the σ-field generated by the sets open in the uniform topology.) It was one of Norbert Wiener's achievements to prove in 1923 that there does exist an assignment of probabilities satisfying these rules, and P_T (the corresponding measure on the Borel sets) is accordingly called Wiener measure. Here we shall take its existence for granted.

Brownian motion, as described by Wiener measure, obeys a transformation law having consequences strange and deep. Suppose that a particle performs a Brownian motion for T units of time, and suppose that, in the function representing its path, we contract the time scale by the factor T and the position scale by the factor $\sqrt{T}$. According to the law in question, the new path will be exactly like that of a particle that has been in Brownian motion for 1 unit of time.

To understand why, let x and y be the old and new paths, so that x lies in $C_0[0, T]$, y lies in $C_0[0, 1]$, and

$$y(t) = \frac{1}{\sqrt{T}} x(tT), \qquad 0 \leqq t \leqq 1. \tag{3}$$

Of course (3) defines a mapping

$$C_0[0, T] \to C_0[0, 1]. \tag{4}$$

The transformation law says that, if x is a random path in $C_0[0, T]$ distributed according to P_T, then y is a random path in $C_0[0, 1]$ distributed according to P_1. (Technically, if ϕ_T is the mapping (4), then $P_1 = P_T\phi_T^{-1}$.) Now, according to (1), the quantity $x(tT)$ is a Gaussian random variable with mean 0 and variance tT. Multiplying a Gaussian random variable by a constant a multiplies its mean by a and its variance by a^2, and the new variable is also Gaussian. Therefore the distribution of $y(t)$ as defined by (3) follows the Gaussian curve with mean 0 and variance t (since $T^{-\frac{1}{2}} \cdot 0 = 0$, $(T^{-\frac{1}{2}})^2 \cdot tT = t$), the first requirement for Brownian motion. Contracting time by the factor T leads from a path over $[0, T]$ to a path over $[0, 1]$, and the vertical rescaling by $1/\sqrt{T}$ makes the variances work out right. Moreover, x has (over disjoint intervals) independent increments, and it is intuitively clear that monotone changes of the time and position scales cannot convert independent increments into dependent ones. So the transformation (3) must preserve the other property of Brownian motion, that of independent increments. This argument, which makes the transformation law plausible, can be converted into a complete proof.

By means of the transformation defined by (3), it is possible to see that, whatever positive values ε and K may have, a Brownian path over $[0, 1]$ will with probability exceeding $1 - \varepsilon$ have somewhere a chord with slope exceeding K. The trick is this: We want, over $[0, 1]$, a Brownian path y with a steep chord. We obtain it not directly, but by applying the transformation (3) to a Brownian path x over $[0, T]$ with T suitably chosen. Choose T so large that x will, with probability exceeding $1 - \varepsilon$, have somewhere a chord with slope exceeding, say, 1. Such a T exists because even the most miraculous event will happen in the long run (the monkeys at the typewriters), and the occurrence of a chord with slope exceeding 1 is a modest miracle indeed. At the same time, choose T to exceed K^2. If x has a chord with slope exceeding 1, and if x and y are related by (3), then y has a chord with slope exceeding $\sqrt{T}$, which in turn exceeds K.

Since ε may be taken arbitrarily small and K arbitrarily large, a Brownian path over $[0, 1]$ must with probability 1 have chords with arbitrarily great slope. There must also be chords with arbitrarily large negative slope, and in fact, chords (very short ones) with extreme slopes are dense along the path. In rigorous and more elaborate form, these arguments show that, if A is the set of paths in $C_0[0, 1]$ of unbounded variation, then $P_1(A) = 1$. A path of unbounded variation represents the motion of a particle that in its wanderings back and forth travels an infinite distance, and at this point physicists lose interest because of their obsession with reality. The fact is mathematically interesting, however, and so is the fact that $P_1(A) = 1$ if A is the set of functions in $C_0[0, 1]$ that are nowhere differentiable. Constructing a

continuous, nowhere differentiable function is difficult, but drawing an element from $C_0[0,1]$ randomly according to P_1 produces such a function with probability 1.

In what follows we shall be mainly concerned with sets that correspond more closely with reality. Although Sections 2 and 3 will involve the transformation (3) and T's that exceed 1, for the rest of this section we shall take $T = 1$. We shall need (1) for the case $T = t = 1$:

$$P_1[x: \alpha \leqq x(1) \leqq \beta] = \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-u^2/2}\, du. \tag{5}$$

Suppose $\alpha \geqq 0$ and consider the event $[x: \max x(t) \geqq \alpha]$ that the particle achieves the height α at some time t with $0 \leqq t \leqq 1$. First,

$$P_1[x: \max x(t) \geqq \alpha] = P_1[x: \max x(t) \geqq \alpha \text{ and } x(1) \geqq \alpha] + P_1[x: \max x(t) \geqq \alpha \text{ and } x(1) < \alpha].$$

The two probabilities on the right here can be proved equal, roughly because once the particle achieves the height α it is as likely, in the absence of drift, to wander upward and finish above α at time 1 as to wander downward and finish below α. Thus

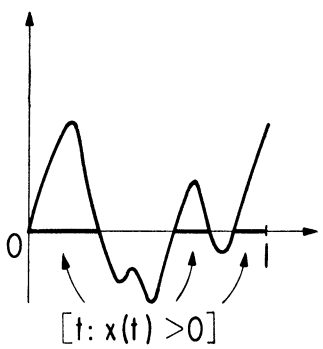
$$P_1[x: \max x(t) \geqq \alpha] = 2P_1[x: \max x(t) \geqq \alpha \text{ and } x(1) \geqq \alpha].$$

Since the condition $\max x(t) \geqq \alpha$ is superfluous in the presence of the condition $x(1) \geqq \alpha$, the right side here is $2P_1[x: x(1) \geqq \alpha]$, and (5) with $\alpha \geqq 0$ and $\beta = \infty$ now implies

$$P_1[x: \max x(t) \geqq \alpha] = \frac{2}{\sqrt{2\pi}} \int_\alpha^\infty e^{-u^2/2}\, du. \tag{6}$$

Thus we have the distribution of the greatest positive excursion.

FIGURE 4



Although to make it rigorous requires some effort, this derivation of (6) has an intuitive appeal. The next result will be stated without any proof, and like many

ex cathedra assertions, it runs counter to intuition. Consider the set $[t: x(t) > 0]$ of time points t, $0 \leqq t \leqq 1$, for which the particle is above 0. This set is a union of intervals (infinitely many, contrary to Figure 4). Denote by bars the Lebesgue measure of this set, the sum of the lengths of the constituent intervals: $\left|[t: x(t) > 0]\right|$. The distribution of this quantity, the total time spent above 0, is given by

$$P_1[x: \alpha \leqq \left|[t: x(t) > 0]\right| \leqq \beta] = \frac{1}{\pi} \int_\alpha^\beta \frac{du}{\sqrt{u(1-u)}} \tag{7}$$

for $0 \leqq \alpha \leqq \beta \leqq 1$. This is Paul Lévy's arc sine law, so called because carrying out the integration leads to the arc sine function.

FIGURE 5

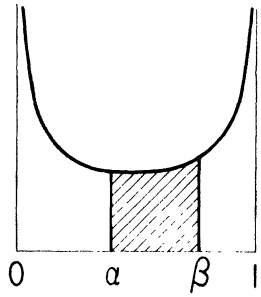
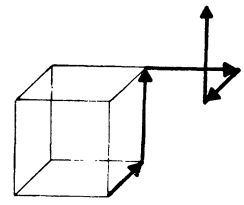


Figure 5 shows the shape of the density, the area of the shaded region representing the right side of (7). The curve is U-shaped, so that if the length $\beta - \alpha$ of the interval is fixed, the probability of $\alpha \leqq \left|[t: x(t) > 0]\right| \leqq \beta$ grows as the interval nears 0 or 1, being smallest when the interval is centered on $\frac{1}{2}$. This is odd because the time spent above 0 has mean $\frac{1}{2}$ by symmetry, and ordinarily values near the mean of a random quantity are more likely to occur than are values far removed from the mean, whereas here the situation is just the opposite.

For general accounts of Brownian motion, see [4] and [7].

2. Random walk. Imagine a particle moving about at random on the nodes of a cubic lattice. The particle can move in any of six directions (north, south, east, west, up, down) to an adjacent node. The direction is determined by the roll of a balanced die, the particle moves to the next node, and the die is rolled once more to determine the direction of the next move, and so on. Figure 6 shows five steps of

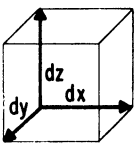
FIGURE 6



such a random walk, together with one of the cells of the cubic lattice. The figure is in

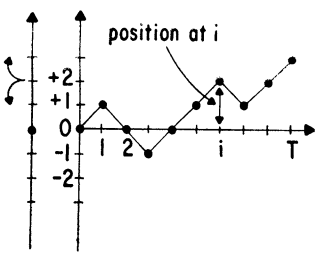
the spirit of a venerable vector analysis book which began a proof of Gauss's theorem by enjoining the reader to consider "an infinitesimal element of volume of dimensions dx, dy, and dz." This injunction was accompanied by a nicely labelled diagram like Figure 7, which was said to show such an infinitesimal element of volume "much enlarged." Well, Figure 6 is much enlarged too, and if the cubes of the lattice are really very small and the particle moves very rapidly from node to node it is natural to expect the motion to approximate Brownian motion.

FIGURE 7



We shall explore a one-dimensional version of this idea. Consider a vertical axis with the integer points 0, ± 1, $\pm 2, \cdots$ marked off on it. We start at 0, toss a coin, and move upward one unit if the coin falls heads and downward one unit if the coin falls tails. In the new position ($+1$ or -1), we toss the coin again and move up or down one unit according as it falls heads or tails, and we continue this way for T steps, T being here an integer. If we take one unit of time to execute each step of this random walk and proceed at a uniform rate from one node to the next, our progress is described by a function like that in Figure 8, a polygonal path whose height over i is the position at i — that is, the position after the ith step. Of the 2^T such paths, each has probability 2^{-T}. (Various aspects of random walk are discussed in [3].)

FIGURE 8



The path can also be viewed as describing the fluctuations in a gambler's fortune. The position on the vertical axis represents the gambler's fortune (relative to his initial capital, so that he starts conventionally at 0), and it moves up or down one unit — say one pound — according as he wins or loses the next play.

The random walk path has some of the properties of a Brownian motion path over $[0, T]$. In the first place, for integers with $i < i' < j < j'$, the displacements undergone over the time intervals $[i, i']$ and $[j, j']$ are independent because they depend on disjoint sets of tosses and the tosses are assumed independent (the coin has no memory). Thus the path has essentially independent increments (for intervals

with nonintegral endpoints the increments can be slightly dependent). The distance moved in one step has mean

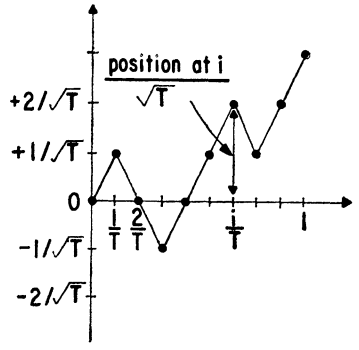
$$(+1)\tfrac{1}{2} + (-1)\tfrac{1}{2} = 0 \tag{8}$$

and variance

$$(+1)^2\tfrac{1}{2} + (-1)^2\tfrac{1}{2} = 1, \tag{9}$$

and so the position at i has mean 0 and by independence has variance i, another property of Brownian motion (see equation (1)). (For nonintegral t, the position at t has mean 0, but the variance is only approximately t.) Although the polygonal

FIGURE 9



character of of the path is not shared by Brownian motion, contraction of the two scales will make the straight-line segments in Figure 8 disappear in the limit as $T \to \infty$.

Suppose we contract the time scale by a factor T and the vertical scale by a factor $\sqrt{T}$, applying the transformation (3) to pass from Figure 8 to Figure 9. In Figure 8 the segments have length $\sqrt{2}$, whereas in Figure 9 they are very short for large T, having length of the order $1/\sqrt{T}$. If Figure 8 represented a Brownian motion path over $[0, T]$, then, as explained in Section 1, Figure 9 would represent a Brownian motion path over $[0, 1]$. The transformation (3) leaves invariant those characteristics (means, variances, independence of increments) the original path shares with Brownian motion and tends to mask those characteristics (piecewise linearity) it does not share. Thus we can hope that the curve in Figure 9 will be very like a Brownian motion path for large T. And indeed, it is true that

$$\text{Prob}[\text{path} \in A] \to P_1(A) \qquad (T \to \infty) \tag{10}'$$

for subsets A of the space $C_0[0, 1]$, where $P_1(A)$ is Wiener measure. There are 2^T paths like the one in Figure 9, and Prob [path $\in A$] is 2^{-T} times the number of them that lie in A.

For an illustration of this theorem, suppose the A in (10) is the set $[x: \alpha \leqq x(1) \leqq \beta]$ of paths in $C_0[0, 1]$ that over the point $t = 1$ have a height between α and β. Since the height over $t = 1$ in Figure 9 is $1/\sqrt{T}$ times the position at T in the original

random walk, (10) and (5) together imply

$$\text{(11)} \qquad \text{Prob}\left[\alpha \leqq \frac{\text{position at } T}{\sqrt{T}} \leqq \beta\right] \to \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-u^2/2}\, du.$$

This is the classical DeMoivre-Laplace central limit theorem for Bernoulli trials. It describes the position after a large number of steps in a random walk, or the gambler's fortune at the end of an evening's play of T ventures. If $-\alpha = \beta = .9$, the limit in (11) is about .6. If $T = 100$, the gambler thus has probability approximately .6 of ending the evening within $.9 \times \sqrt{100} = 9$ pounds of his initial capital.

Suppose now that A is the set in (6), the set of paths in $C_0[0, 1]$ having somewhere a height at least α (here $\alpha \geqq 0$). The path in Figure 9 lies in A if at some time during the evening's play the gambler's fortune is at least $\alpha\sqrt{T}$ pounds above his initial capital, and by (10), the probability of this converges to the right side of (6). For $\alpha = 1.7$, the value of this limit is about .1. With $T = 100$, this gives an approximate probability of .1 that the gambler will have been at least $1.7 \times \sqrt{100} = 17$ pounds ahead at the time he should have quit.

Finally, suppose A is the set $[x: \alpha \leqq \big|[t: x(t) > 0]\big| \leqq \beta]$ in (7). During the evening the gambler is ahead a certain fraction of the time; if the curve in Figure 9 represents the history of his fortunes, it belongs to the set A if and only if this fraction lies between α and β. The chance of this event is by (10) and (7) about equal to the area of the shaded region in Figure 5. If we compute the areas, the chance the gambler is ahead between 45% and 55% of the time turns out to be only about .06, whereas the chance he is ahead more than 90% of the time is about .2. In one evening in five the gambler will thus be ahead more than 90% of that evening's play. By symmetry, in one evening in five the gambler will be ahead less than 10% of that evening's play. To convince him in the first [second] case that his experience is due merely to chance and not to his being Fortune's favorite [Fortune's fool] will be difficult [impossible].

We have applied (10) to three interesting sets A. If A is the set of functions in $C_0[0, 1]$ of unbounded variation, then $P_1(A) = 1$, as explained in Section 1, while $\text{Prob}[\text{path} \in A] = 0$ because the curve in Figure 9 is visibly of bounded variation. Thus (10) fails for certain subsets A of $C_0[0, 1]$. The mathematical fact is that (10) holds for every set (Borel set) A whose boundary ∂A (boundary in the sense of the uniform topology) satisfies $P_1(\partial A) = 0$—a condition which holds in our three applications but not if A is the set of functions of unbounded variation. A complete proof of this theorem uses a combination of probability theory and functional analysis; the details can be found in [**1**].

3. Prime divisors. According to the fundamental theorem of arithmetic, each integer has a factorization into primes, a factorization unique except for order (see [**5**], for example). Let $f(n)$ be the number of distinct primes in the factorization of n; we do not count multiplicity: $f(3^4 \cdot 5^2)$ is 2, not 6. The table shows some values of the

n	2	3	4	5	6	7	⋯	29	30	31	⋯	209	210	211	⋯
$f(n)$	1	1	1	1	2	1	⋯	1	3	1	⋯	2	4	1	⋯

function f. It rises slowly. The smallest n's with respective f-values 2, 3, and 4 are $2 \cdot 3 = 6$, $2 \cdot 3 \cdot 5 = 30$, and $2 \cdot 3 \cdot 5 \cdot 7 = 210$. The fact that there are infinitely many primes implies, however, that f assumes arbitrarily large values; since $f(p) = 1$ for prime p, the same fact implies that f infinitely often drops back to 1.

Since f varies in this irregular fashion, it is natural to ask after its average behavior. For example, it can be shown that

$$\frac{1}{N}\sum_{n=1}^{N} f(n) \approx \log\log N \tag{12}$$

(see the remarks following (17) below). Since $\log\log 10^{70} \approx 5$, the typical integer under 10^{70} has a mere five prime divisors. More delicate questions concern the distribution of f. If S is a set of positive integers, let $\mathbf{P}_N(S)$ be the fraction among the integers $1, 2, \cdots, N$ that lie in S:

$$\mathbf{P}_N(S) = \frac{1}{N} \times \#[n: 1 \leqq n \leqq N \text{ and } n \in S]. \tag{13}$$

The problem is to get information about quantities like $\mathbf{P}_N[n: a \leqq f(n) \leqq b]$.

Now (13) can be viewed as a probability: We draw an integer at random from the range $1 \leqq n \leqq N$, and $\mathbf{P}_N(S)$ is the probability that it will lie in S. That $\mathbf{P}_N[n: a \leqq f(n) \leqq b]$ can be viewed as a probability does not by itself ensure (this may be difficult to credit) that probability theory will help in the evaluation. It does in fact help because the notion of independence can be brought to bear. If $\delta_p(n)$ is 1 or 0 according as the prime p divides n or not, then $f(n) = \Sigma_p \delta_p(n)$. We can understand the distribution of $f(n)$ if we understand the joint behavior of the $\delta_p(n)$ as random quantities.

The number of multiples of p up to N is the integral part $[N/p]$ of N/p. The probability that $\delta_p(n) = 1$, or that $p \mid n$, is thus

$$\mathbf{P}_N[n: p \mid n] = \frac{1}{N}\left[\frac{N}{p}\right] \approx \frac{1}{p}. \tag{14}$$

The approximation here is good for large N: since $[N/p]$ differs from N/p by less than 1, the error in (14) is less than $1/N$. The formula (14) reflects the fact that p divides every pth integer, and it in no way requires that p be prime.

The fundamental theorem of arithmetic implies that, if integers a and b are relatively prime (share no prime factors), then they individually divide n if and only if their product ab divides n. This fact is well illustrated by the use Turing is said to have made of it. The sprocket wheel of his bicycle had a faulty tooth and the chain a faulty link, and unless he was pedalling very fast when the faulty parts meshed, the chain would fall off. So he counted the number, say a, of teeth on the wheel and the

number, say b, of links on the chain and found, not to his surprise, that a and b were relatively prime. Between successive meetings of the bad tooth and link the sprocket wheel would in consequence go through b cycles, as the chain went through a cycles. Turing is said to have pedalled along counting, on every bth cycle of the sprocket wheel giving the burst of speed necessary to carry him past the danger point.

As a special case of this fact, distinct primes p and q individually divide n if and only if pq does. By this and by (14) with pq in place of p,

$$\mathbf{P}_N[n: p \mid n \text{ and } q \mid n] = \mathbf{P}_N[n: pq \mid n] = \frac{1}{N}\left[\frac{N}{pq}\right] \approx \frac{1}{pq} = \frac{1}{p} \cdot \frac{1}{q}.$$

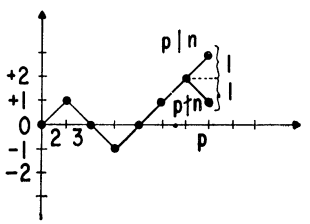
Since by (14) the factors $1/p$ and $1/q$ respectively approximate $\mathbf{P}_N[n: p \mid n]$ and $\mathbf{P}_N[n: q \mid n]$ if N is large, we arrive at

$$\mathbf{P}_N[n: p \mid n \text{ and } q \mid n] \approx \mathbf{P}_N[n: p \mid n]\,\mathbf{P}_N[n: q \mid n]. \tag{15}$$

Thus the events $[n: p \mid n]$ and $[n: q \mid n]$ approximately satisfy the definition of independence if n is random, $1 \leqq n \leqq N$, with N large. There is an extension of (15) from two primes to three or more.

We can use this fact to construct a kind of random walk path containing information about the prime factorization of n and in particular about $f(n)$. We draw an integer n at random from among $1, 2, \cdots, N$. On a vertical axis with the integer points marked off on it, we start at 0 and go up one unit if $2 \mid n$ and down one unit if $2 \nmid n$. From our new position ($+1$ or -1), we go up one unit if $3 \mid n$ and down one unit if $3 \nmid n$. We proceed in this way, examining each prime in succession. Figure 10 describes this factorization random walk in the same way that Figure 8 describes the coin-tossing random walk. Each number on the time axis is the prime corresponding to that step in the random walk. We consider later how long to continue the walk.

FIGURE 10



Since n is random, this path is random. But since the randomness is all in the drawing of n before the walk starts, the factorization random walk may seem less random than the coin-tossing random walk. This is an illusion. We may imagine tossing the coin T times in advance of the walk, recording the sequence of heads and tails, and only then performing the corresponding walk. Since we would see its whole history on record before setting out, the walk would be very dull. So imagine a friend who tosses the coin T times and records the results in advance of the journey, and imagine that, rather than show us the record all at once, he instead reveals the

outcomes to us one by one as we execute the walk. This restores the suspense. For the factorization random walk, we can imagine a friend who draws n at random, $1 \leqq n \leqq N$, factors n into primes, and at each step of the walk reveals to us whether or not the corresponding p divides n.

The increment of the random path in Figure 10 over an interval depends on how many in the corresponding set of primes divide n. Increments over disjoint intervals depend on disjoint sets of primes and hence by (15)—or by (15) together with its extension to three or more primes—the increments will be approximately independent if N is large. Unlike Brownian motion, however, the factorization random walk has a strong downward drift. By (14), the chance of going downward at the step corresponding to p is about $1 - 1/p$, which is almost 1 for large p. The remedy is to move up a distance $1 - 1/p$ if $p \mid n$ and to move down only a distance $1/p$ if $p \nmid n$. The expected distance moved is now

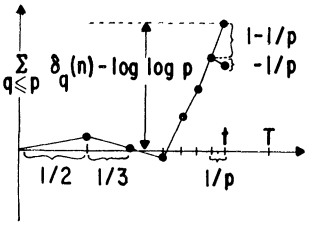
$$(1 - p^{-1})\mathbf{P}_N[n: p \mid n] + (-p^{-1})\mathbf{P}_N[n: p \nmid n],$$

which by (14) is approximately

$$\left(1 - \frac{1}{p}\right)\frac{1}{p} + \left(-\frac{1}{p}\right)\left(1 - \frac{1}{p}\right) = 0.$$

This corresponds with (8), an equation which shows that the coin-tossing random walk has no drift.

FIGURE 11



Since the mean distance moved at the step corresponding to p is approximately 0, the variance is approximately $(1 - p^{-1})^2\mathbf{P}_N[n: p \mid n] + (-p)^2\mathbf{P}_N[n: p \nmid n]$, which by (14) is in turn approximately

$$\left(1 - \frac{1}{p}\right)^2\frac{1}{p} + \left(-\frac{1}{p}\right)^2\left(1 - \frac{1}{p}\right) = \frac{1}{p}\left(1 - \frac{1}{p}\right) \approx \frac{1}{p}.$$

The distance moved thus tends to be very small for large p, in contrast with the coin-tossing random walk, which by (9) proceeds with vigor ever undiminished. The remedy this time is to spend only an amount of time $1/p$ executing the step corresponding to p. With these two modifications, the path is as in Figure 11.

To recapitulate, the time interval corresponding to the prime p has length $1/p$. Over this interval, the path rises an amount $\delta_p(n) - 1/p$; that is, it rises $1 - 1/p$ if $p \mid n$ (the probability of this is approximately $1/p$) and it rises $0 - 1/p$ if $p \nmid n$ (the probability of this is approximately $1 - 1/p$).

The point t in Figure 11 (the right endpoint of the interval corresponding to p) is $\Sigma_{q\leqq p}\, 1/q$ (summation over primes q not exceeding p). The distance moved in the step corresponding to q has variance about $1/q$, and hence by the approximate independence of the steps ((15) again), the variance of the position at this time t is approximately $\Sigma_{q\leqq p}\, 1/q$, or t itself. The above adjustment of the factorization random walk has thus not only eliminated the drift, it has so adjusted the time scale that the variances are about what they are for the coin-tossing random walk and for Brownian motion.

It can be shown that

$$\sum_{q\leqq u} \frac{1}{q} \approx \log\log u \tag{16}$$

for large u (the two expressions go to infinity with u and their difference remains bounded; see [5, p. 351]). That the sum in (16), instead of increasing in some erratic fashion, is asymptotic to a standard function like log log u is inessential to what follows; but the formulas become simpler (and remain valid) if at each occurrence of the sum we substitute the right side of (16).

Thus the t in Figure 11 is essentially log log p and the height $\Sigma_{q\leqq p}(\delta_q(n) - 1/q)$ of the curve over t is approximately

$$\sum_{q\leqq p} \delta_q(n) - \log\log p. \tag{17}$$

Now n has $\Sigma_{q\leqq p}\, \delta_q(n)$ prime divisors that do not exceed p, and we normalize this quantity by subtracting away the value log log p it has for a "typical" n. (If n is random, $1 \leqq n \leqq N$, then $\Sigma_{q\leqq p}\, \delta_q(n)$ has by (14) and (16) a mean of about log log p; this is where (12) comes from.) The factorization random walk is a record of these differences (17). We continue the walk until each $p \leqq N$ has been dealt with, and the corresponding point on the time axis is $T = \Sigma_{p\leqq N}\, 1/p \approx \log\log N$.

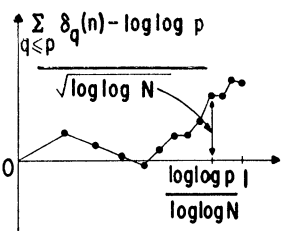


FIGURE 12

The random path now resembles a coin-tossing path in that the increments are almost independent for large N, there is essentially no drift, and the variances are about right. As in the coin-tossing case, rescaling will lead in the limit ($N \to \infty$) to Brownian motion. To send T to the point 1, we contract the horizontal scale by a factor $T = \log\log N$, and, again as in the coin-tossing case and for the same reasons, we contract the vertical scale by the square root of this, applying the transformation

(3). The point t in Figure 11 goes to $\log\log p/\log\log N$, and the path is that shown in Figure 12.

Since the path depends on n and N, denote it $\text{path}_N(n)$. Since n is random ($1 \leqq n \leqq N$), so is the path, and the chance that it lies in a given subset A of $C_0[0,1]$ is $\mathbf{P}_N[n\colon \text{path}_N(n) \in A]$. The theorem linking primes with Brownian motion is this: If A is a subset (Borel subset) of $C_0[0,1]$ satisfying $P_1(\partial A) = 0$, then

$$\mathbf{P}_N[n\colon \text{path}_N(n) \in A] \to P_1(A) \qquad (N \to \infty), \tag{18}$$

where $P_1(A)$ is Wiener measure. The proof of (18) uses a combination of probability theory, functional analysis, and number theory. The theorem is given implicitly in [8, p. 122], explicitly in a manuscript version of [1] and in a much more general form in [9]. (For general discussions of probability methods in number theory, see [6], [8] and the author's 1973 Wald lectures, to appear in the *Annals of Probability*.)

From Figure 12, a plot of the differences (17) normalized to

$$\Big(\sum_{q \leqq p} \delta_q(n) - \text{og}\log p\Big)/\sqrt{\log\log N}, \tag{19}$$

we can read off arithmetic properties of n, and therefore (18) yields arithmetic limit theorems. Consider the three sets A to which we applied the analogous result (10). The height of the curve in Figure 12 over the time point 1 is (19) with N in place of p; it is the number $f(n)$ of prime factors of n, normalized to

$$(f(n) - \log\log N)/\sqrt{\log\log N}.$$

The greater this is, the more highly composite n is, and the smaller it is, the more "prime-like" n is. With $A = [x\colon \alpha \leqq x(1) \leqq \beta]$, it follows by (18) and (5) that

$$\mathbf{P}_N\left[n\colon \alpha \leqq \frac{f(n) - \log\log N}{\sqrt{\log\log N}} \leqq \beta\right] \to \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-u^2/2}\, du. \tag{20}$$

This is the Erdös-Kac central limit theorem for f. (For an elementary direct proof of (20), see [2].)

For $-\alpha = \beta = .9$, the limit in (20) is about .6, and if $N = 10^{70}$, so that $\log\log N \approx 5$, the double inequality in (20) is approximately the same as $-.9 \leqq (f(n) - 5)/\sqrt{5} \leqq .9$, which in turn is approximately the same as $3 \leqq f(n) \leqq 7$. Thus something like 60% of the integers under 10^{70} have from 3 to 7 prime divisors.

The larger (17) is, the more highly composite n appears to be at that point in the factorization; that is, (17) measures the apparent compositeness of n when it has been tested for divisibility only by primes up to p. The maximum apparent compositeness is measured by

$$\max_{p \leqq N} \Big(\sum_{q \leqq p} \delta_q(n) - \log\log p\Big); \tag{21}$$

since this is $\sqrt{\log\log N}$ times the maximum height of the curve in Figure 12, an application of (18) to the set in (6) gives its approximate distribution. The right side of (6) being about .1 if $\alpha = 1.7$, for about 10% of the integers under 10^{70} does (21) exceed $1.7 \times \sqrt{5} \approx 3.8$.

Let us say that n is *excessive at* p if

$$\sum_{q \leq p} \delta_q(n) > \log\log p; \tag{22}$$

this holds if, with respect to divisibility by primes up to p, n is "more composite," or "less prime-like," than the average integer. And (22) holds exactly when the corresponding point on the curve in Figure 12 is above the axis. The polygonal segment corresponding to p has length $p^{-1}/\log\log N$ when projected on the horizontal axis, and so the amount of time the curve spends above 0 is essentially

$$\frac{1}{\log\log N} \sum \left[\frac{1}{p} : p \leq N \text{ and } \sum_{q \leq p} \delta_q(n) > \log\log p\right], \tag{23}$$

the sum extending over those p at which n is excessive.

If we test n for divisibility by the primes in succession, spending an amount $1/p$ of time on p ($p \leq N$), (23) is the fraction of time we are dealing with a p at which n is excessive. From an application of (18) to the set in (7) it follows that for large N the distribution of (23) approximately follows the density curve in Figure 5. For about 20% of the integers under N the quantity (23) exceeds .9, for about 20% it is less than .1, and for only about 6% does it lie between .45 and .55.

Prime factors exhibit in this respect the same strange behavior coins do. In a way they are even more strange. A quantity perhaps more natural to consider than (23) is

$$\frac{1}{\pi(N)} \times \#[p : p \leq N \text{ and } \sum_{q \leq p} \delta_q(n) > \log\log p], \tag{24}$$

the *number* of p for which n is excessive at p, normalized by division by $\pi(N)$, the total number of primes involved. For N large, of the break points in the polygon in Figure 12 the great majority are very near 1, which has the result that in the limit the distribution of (24) consists of a mass of $\frac{1}{2}$ at 0 and a mass of $\frac{1}{2}$ at 1: If $\varepsilon > 0$ and N exceeds some N_ε, then (24) is less than ε with a probability lying in the range $\frac{1}{2} - \varepsilon$ and $\frac{1}{2} + \varepsilon$ and is greater than $1 - \varepsilon$ with a probability lying in the same range. Thus practically all integers are excessive either at practically all primes or at practically none.

The 1972 Rouse Ball Lecture, given while the author was a Guggenheim Fellow, visiting Peterhouse and the Statistical Laboratory of the University of Cambridge. It appeared in somewhat different form in *Eureka*, the Journal of the Archimedeans, the Cambridge University Mathematical Society.

References

1. Patrick Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.
2. ———, On the central limit theorem for the prime divisor function, this MONTHLY, 76 (1969) 132–139.
3. William Feller, An Introduction to Probability Theory and Its Applications, vol. I, 3rd ed., Wiley, New York, 1968.
4. David Freedman, Brownian Motion and Diffusion, Holden-Day, San Francisco, 1971.
5. G. H. Hardy and E. M .Wright, An Introduction to the Theory of Numbers, 4th ed., Clarendon Press, Oxford, 1960.
6. M. Kac, Statistical Independence in Probability, Analysis and Number Theory, Carus Math. Monogr. 12. MAA, Wiley, New York, 1959.
7. Samuel Karlin, A First Course in Stochastic Processes, Academic Press, New York, 1966.
8. J. Kubilius, Probabilistic Methods in the Theory of Numbers, 2nd ed. (1962). Vilna: Gosudarstv. Izdat. Politič. i Naučn. Lit. Litovsk. SSR. (English translation 1964. Amer. Math. Soc. Transl. of Math. Monographs, Volume 11.)
9. Walter Philipp, Arithmetic functions and Brownian motion, Proc. Symp. Pure Math., vol. 24, AMS, 1973.

CORRECTION TO "UNIQUE FACTORIZATION DOMAINS"

P. M. COHN, Bedford College, University of London

The statement "Any Noetherian UFD is a Dedekind domain" (this MONTHLY, 80 (1973) 1–18) should be omitted.

The assertion is of course well known to be false; a correct statement would be: A Dedekind domain is a UFD if and only if it is a principal ideal domain.

I am indebted to Professor J. H. Hays for drawing my attention to this error.

CORRECTION TO "A HISTORY OF THE PRIME NUMBER THEOREM"

L. J. GOLDSTEIN, University of Maryland

In my paper, [this MONTHLY, 80 (June-July, 1973) 599–615] I asserted that the sieve of Eratosthenes was known to the ancient Greeks and, in fact, appeared in Euclid. It has been pointed out to me by Professor J. Albree that although the sieve was known since approximately the time of Euclid, it does not appear in the *Elements*. The author regrets the error.