24 PROBABILITY

MARK KAC • September 1964

A secretary has typed 10 letters and addressed 10 envelopes. If she now puts the letters in the envelopes entirely at random (without looking at the addresses), what is the probability that not a single letter will wind up in its correct envelope? It may surprise the reader to learn that the probability is better than one chance in three: more specifically, it is almost 1/2.71828.... (This famous number 2.71828..., or e, the base of the natural logarithms, turns out to be an important one in the theory of probability and comes up again and again, as we shall see.)

The method used to solve the problem is called combinatorial analysis. An older and more familiar example of problems in combinatorial analysis is: What is the probability of drawing a flush in a single deal of five cards from a deck of 52? Combinatorial analysis has more profound and more practical applications, of course, than estimating the chances of poker hands or answering amusing questions about the hypothetical behavior of absentminded secretaries. It has become an extremely useful branch of mathematics. But its principles are best illustrated by simple examples. Let us work out the poker problem in detail so that we can perceive some of its probabilistic implications.

Pierre Simon de Laplace (1749–1827) based an entire theory of probability on combinatorial analysis by defining probability as $p = n/N$. This expression states that the probability of an event is the ratio of the number of ways in which the event can be realized (n) to the total number of possible events (N), provided that all the possible events are equally likely—an important proviso. The probability of a poker flush therefore is the ratio of the number of possible flushes to the total number of possible poker hands. The problem of combinatorial analysis is to calculate both numbers.

Let us start with a simpler case involving more manageable numbers. Given a set of four objects, A, B, C, D, how many subsets, or combinations, of two objects can be made from them? It is easy to answer by simple pairing and counting: there are six possible combinations of size 2, AB, AC, AD, BC, BD, CD. As we go on to larger numbers of objects, however, this process soon becomes all but impossible. We must find shortcuts—ways to make the calculations without actually counting. (Combinatorial analysis is sometimes called "counting without counting.")

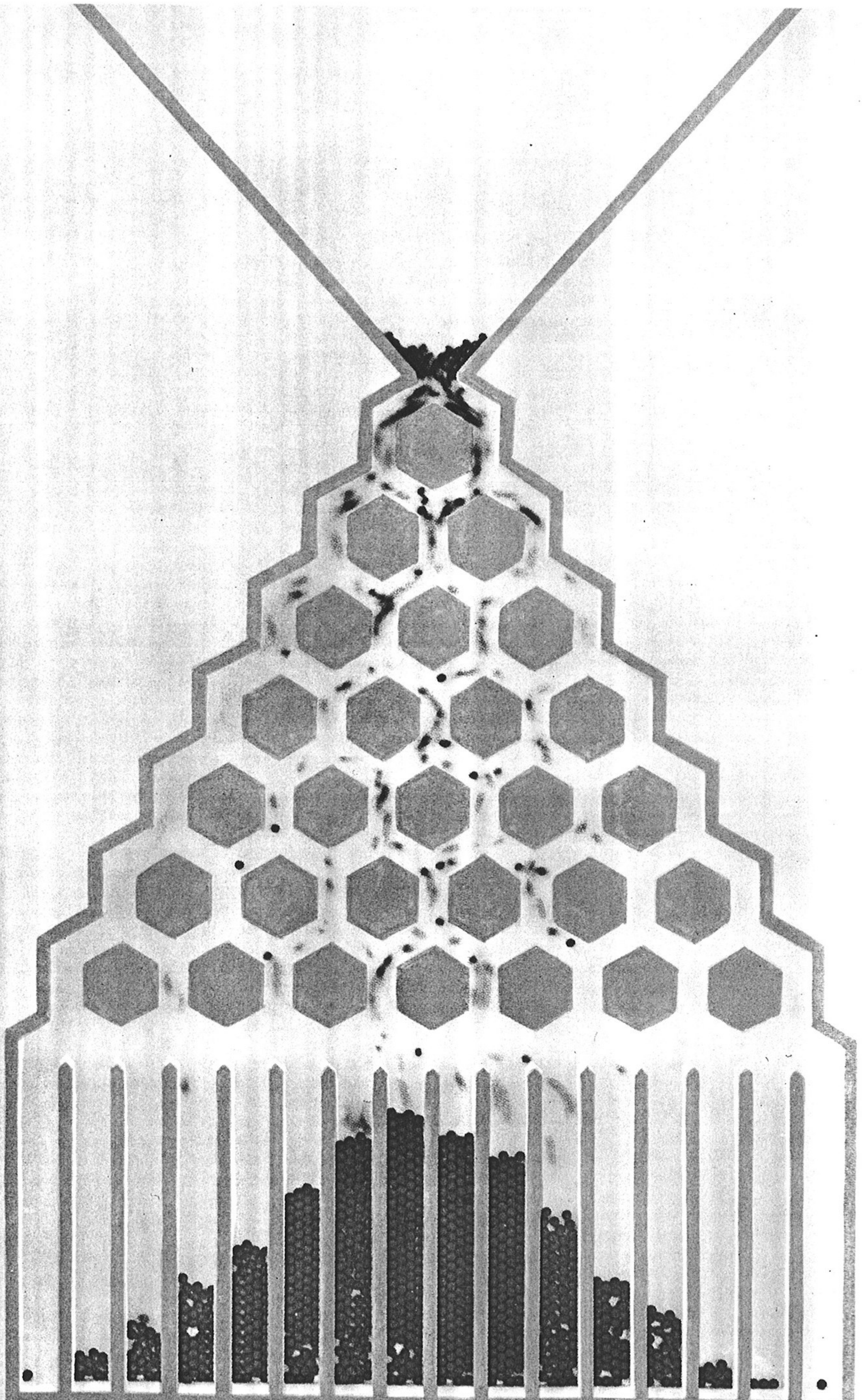
Suppose we add a fifth object and consider how many pairs can be formed from the five. It is apparent that the new object, E, adds just four to the total of possible pairs, because it can combine with each of the other four. So the total is 6 + 4, or 10, possible pairs. To put it in the conventional symbols of combinatorial analysis, we have $C(5,2) = C(4,2) + C(4,1)$. C represents the number of combinations, and the numbers in parentheses stand respectively for the total number of objects and the number in each subset: for instance, $C(5,2)$ means the combinations of five objects taken two at a time. To calculate on the same principle the number of combinations of four objects that could be made out of a total of 10 objects, we could write $C(10,4) = C(9,4) + C(9,3)$ and then continue the reduction to smaller and smaller numbers until we finally computed the answer by simple addition of all the numbers. In practice what we actually do in such a case is to build the C's from the ground up (the bookkeeping is easier).

The whole scheme is conveniently summarized in a handy table known as Pascal's triangle after Blaise Pascal (1623–1662), one of the founders of the theory of probability. The triangle is made up of the coefficients of the binomial expansion, each successive row representing the next higher power [*see top illustration on page 167*]. Each number in the table is the sum of the two numbers to the right and the left of it in the row above. The number of combinations for any set of objects can be read from left to right across a row. For example, the fourth row describes the possible combinations when the total number of objects is four: reading from the left, we have first the number 1, for the "empty set" (containing no objects); then 4, the number of subsets containing one object; then 6, the number of possible combinations of two objects; then 4, the number of three-object combinations, and finally 1 for the "full set" of four objects. With this table, to find the number of quadruplets that can be formed from a total of 10 objects one goes to the 10th row, reads across five steps to the right and finds the answer: 210.

Even the Pascal triangle becomes inconvenient when it has to be extended to large numbers such as are involved in our poker problem. Fortunately the pioneers of the theory of probability were able to work out and prove a simple general formula. This now familiar formula, in which (n,r) means n objects taken r at a time and "!" is the symbol meaning "factorial," is

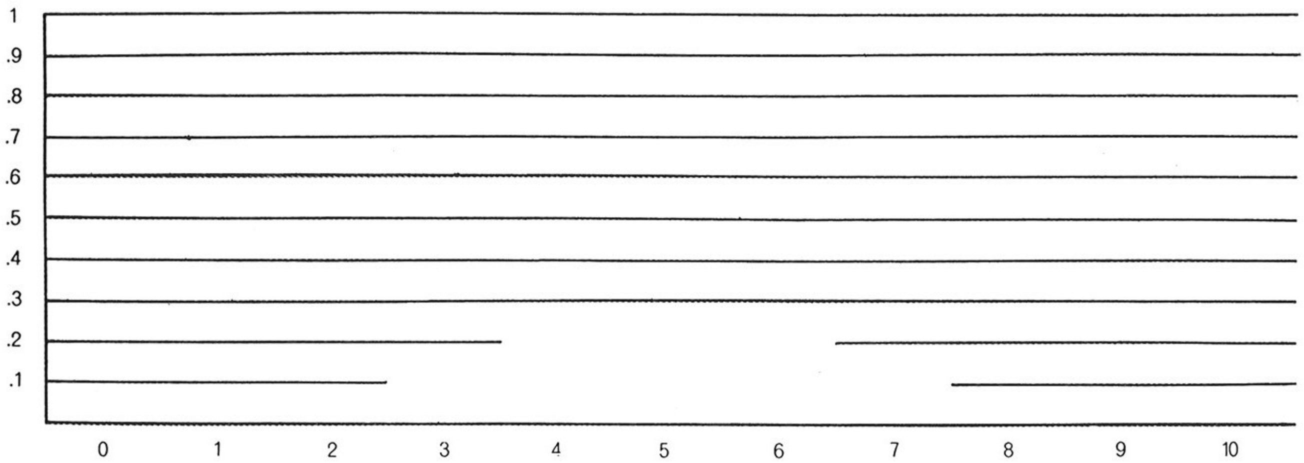
$$C(n,r) = \frac{n!}{r!(n-r)!}\,.$$

In the case of $C(10,4)$ the formula—simplified by dividing both numerator



1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1
1 8 28 56 70 56 28 8 1
1 9 36 84 126 126 84 36 9 1
1 10 45 120 210 252 210 120 45 10 1

PASCAL'S TRIANGLE, an aid to calculating probabilities, is made up of the coefficients of the binomial expansion. Each number is the sum of the two numbers immediately above it. Some of its characteristics and applications are discussed in the text.



PROBABILITY OF HEADS in 10 tosses of a coin produces a histogram that is reminiscent of the normal distribution. As the 10th row of Pascal's triangle shows, there are 210 possible sequences containing exactly four heads out of a total of 1,024 possible sequences of heads in 10 tosses. Thus the chance of four heads is about 21 percent. Roughly the chances of zero, one, two, three, four and five heads in 10 tosses (*horizontal scale*) are respectively .001, .01, .045, .12, .21 and .25 (*vertical scale*). The probabilities and the bars representing them occur in descending order for six through 10 heads in 10 tosses. On the same scale a histogram for 10,000 tosses would be much wider and lower, and would have to be rescaled to bring out its relation to the normal curve.

and denominator by $(n-r)!$—becomes

$$C(10,4)=\frac{10\cdot 9\cdot 8\cdot 7}{1\cdot 2\cdot 3\cdot 4}=210.$$

Now it is not difficult to compute the probability of a poker flush. There are $C(13,5)$ possible flushes in each suit, a total of $4C(13,5)$ in the four suits. The total number of possible poker hands is $C(52,5)$. Hence the probability of getting a flush out of all the hands that might be dealt is

$$\frac{4C(13,5)}{C(52,5)}=\frac{4\,\frac{13\cdot 12\cdot 11\cdot 10\cdot 9}{1\cdot 2\cdot 3\cdot 4\cdot 5}}{\frac{52\cdot 51\cdot 50\cdot 49\cdot 48}{1\cdot 2\cdot 3\cdot 4\cdot 5}}=\frac{33}{16{,}600}.$$

This comes out to about two chances in 1,000 of drawing a flush of any kind in a five-card deal from a full deck.

Let us proceed to investigate probability further by the classic device of coin-tossing. Suppose I toss a coin 10 times; what is the probability that in the 10 throws I shall get exactly four

PROBABILITY DEMONSTRATOR shown on the opposite page mechanically produces an approximation to the bell-shaped "normal," or Gaussian distribution. The little red balls rolling from the reservoir at the top pass an array of hexagonal obstacles and collect in receptacles at the bottom. At each obstacle the probability ought in theory to be one-half that a ball will go to the right and one-half that it will go to the left. Thus the balls tend to distribute themselves according to the proportions of Pascal's triangle, shown in the top illustration above. In the photograph the balls falling through the array of obstacles are blurred because of their motion. The balls have not produced the full distribution curve because some are still moving through the channels. The apparatus is known as a Galton Board after Sir Francis Galton, who constructed the first one. The version shown here is based on a design by the Science Materials Center, patented under the name "Hexstat."

heads? Looking at the 10th row of the Pascal triangle, we see that the possible sequences of heads and/or tails for 10 tosses add up to a total of 1,024. In this total there are 210 sequences containing exactly four heads. Therefore, if the coin-tossing is "honest," in the sense that all the 1,024 possible outcomes are equally likely, the probability of just four heads in 10 throws is 210/1,024, or roughly 21 percent.

The sum of all the entries in any given row (numbered n) of the Pascal triangle is equal to 2 to the nth power (for example, $1,024 = 2^{10}$). Thus in general the probability of tossing exactly k heads in a sequence of n throws is $C(n,k)/2^n$. Suppose we plot the various probabilities of tossing exactly 0, 1, 2, 3 and so on up to 10 heads in 10 throws in the form of a series of rectangles, with the height of each rectangle representing the probability [*see bottom illustration on preceding page*]. The graph peaks at the center (a probability of 252/1,024 for five heads) and tapers off gradually to both sides (down to probabilities of 1/1,024 for no heads and for 10 heads). If we plot the same kind of graph for 10,000 tosses, it becomes much wider and lower: the high point (for 5,000 heads) is not in the neighborhood of 25 percent but only $1/100\sqrt{\pi}$, or approximately .56 percent. (It may seem odd that in increasing the number of tosses we greatly reduce the chances of heads coming up exactly half the time, but the oddity disappears as soon as one realizes that a strict 50–50 division between heads and tails is still only one of the possible outcomes, and with each toss we have increased the total number of possible results.)

Drawn on the basis I have just described, the probability graph for a large number of tosses is so flat that it is hardly distinguishable from a straight line. But by increasing the heights of all the rectangles by a certain factor ($\sqrt{n/2}$) and shrinking the width of the base by the same factor, one can see that the tops of the rectangles trace out a symmetrical curve with the peak in the middle. The larger the number of tosses, the closer this profile comes to a smooth, continuous curve, which is described by the equation

$$y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

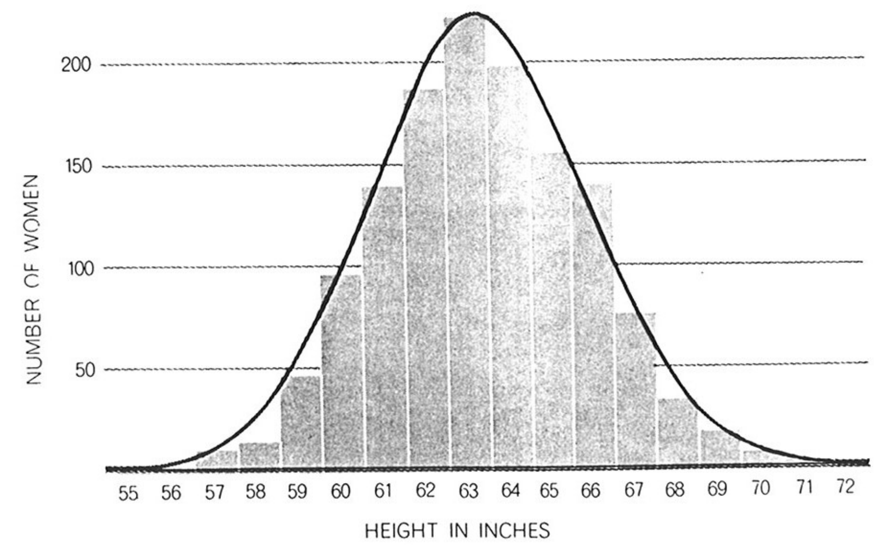
The e is our celebrated number 2.71828..., the base of the natural logarithms. (If a bank were foolish enough to offer interest at the annual rate of 100 percent and were to compound this interest continuously—not just daily, hourly or even every second but every instant—one dollar would grow to $2.71828... at the end of a year.)

The close approach of the probability diagram to a continuous curve with many tosses of a coin illustrates what is called a law of large numbers. If an "honest" coin is tossed hundreds of thousands or millions of times, the distribution of heads in the series of trials, when properly centered and scaled on a graph, will follow almost exactly the curve whose formula I have just given. This curve has become one of the most celebrated in science. Known as the "normal" or "Gaussian" curve, it has been used (with varying degrees of justification) to describe the distribution of heights of men and of women, the sizes of peas, the weights of newborn babies, the velocities of particles in a gas and numerous other properties of the physical and biological worlds.

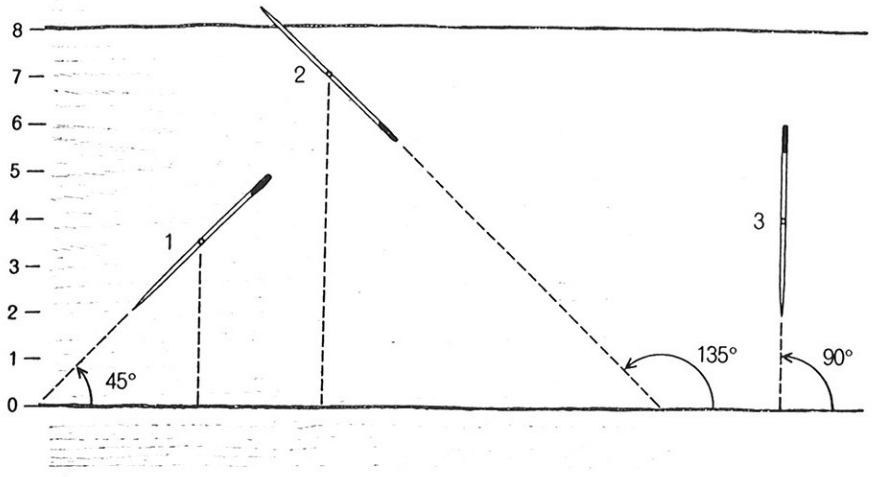
The remarkable connection between coin-tossing and the normal curve was both gratifying and suggestive. It provided one of the main stimuli for the further development of probability theory. It also formed the basis for the "random walk" model of tracing the paths of particles. This in turn solved the mystery of Brownian motion, thus establishing the foundations of modern atomic theory.

Probability today is a cornerstone of all the sciences, and its daughter, the science of statistics, enters into all human activities. How prophetic, in retrospect, are the words of Laplace in his pioneering work *Théorie analytique des probabilités*, published in 1812: "It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge. ... The most important questions of life are, for the most part, really only problems of probability."

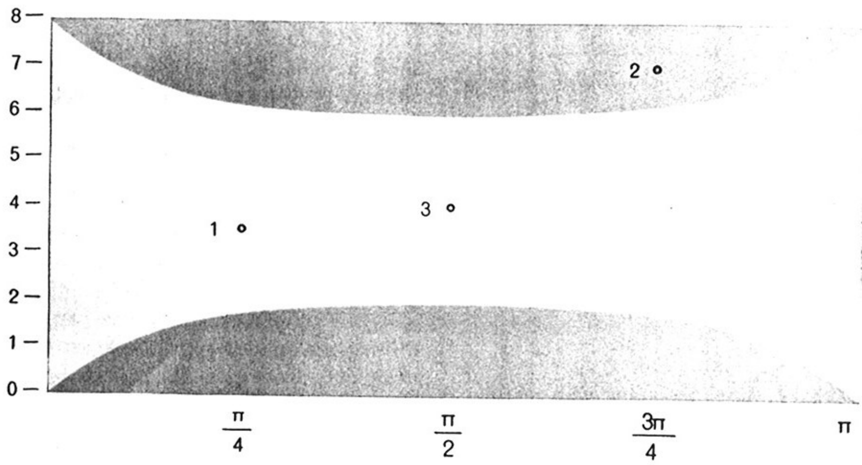
It seems to be a characteristic of "the most important objects of human knowledge" that they generally take a long time to become established as such. After Laplace interest in probability theory declined, and through the rest of the 19th century and the first two decades of the 20th it all but disappeared as a mathematical discipline. Only a few mathematicians went on with the work; among these were the brilliant and original Russian mathematicians P. L. Chebyshëv and his pupil A. A. Markov (which accounts for the strong development of probability theory in the U.S.S.R. today). There were spectacular applications of probability theory to physics, not only by Albert Einstein and Marian Smoluchowski in their solution of the problem of Brownian motion but also by James Clerk Maxwell, Ludwig Boltzmann and Josiah Willard Gibbs in the kinetic theory of gases. At the turn of the century Henri Poincaré and David Hilbert,



HEIGHTS OF WOMEN produce a histogram to which the normal-distribution curve can be fitted. There were 1,375 women in this sample population. The bell-shaped curve conforms to many other empirical distributions found in the physical and biological worlds.



BUFFON NEEDLE PROBLEM involves the probability that a needle shorter than the width of a plank will fall across the crack between two planks. Here each needle is half as long as the plank is wide. The eight units used to measure the plank represent inches.



ABSTRACT DIAGRAM also shows positions of the three needles. The horizontal scale represents the angle of each needle with respect to the bottom edge of the plank. The angle is given in terms of π, which is defined as 180 degrees. The vertical scale is the width of the plank in inches. The three dots are the center points of the needles. Called the "sample space," the rectangle represents all the possible positions in which a needle can fall. The dark colored areas cover all the positions in which a needle lies across a crack.

the two greatest mathematicians of the day, tried to revive interest in probability theory, but in spite of their original and provocative contributions there was remarkably little response.

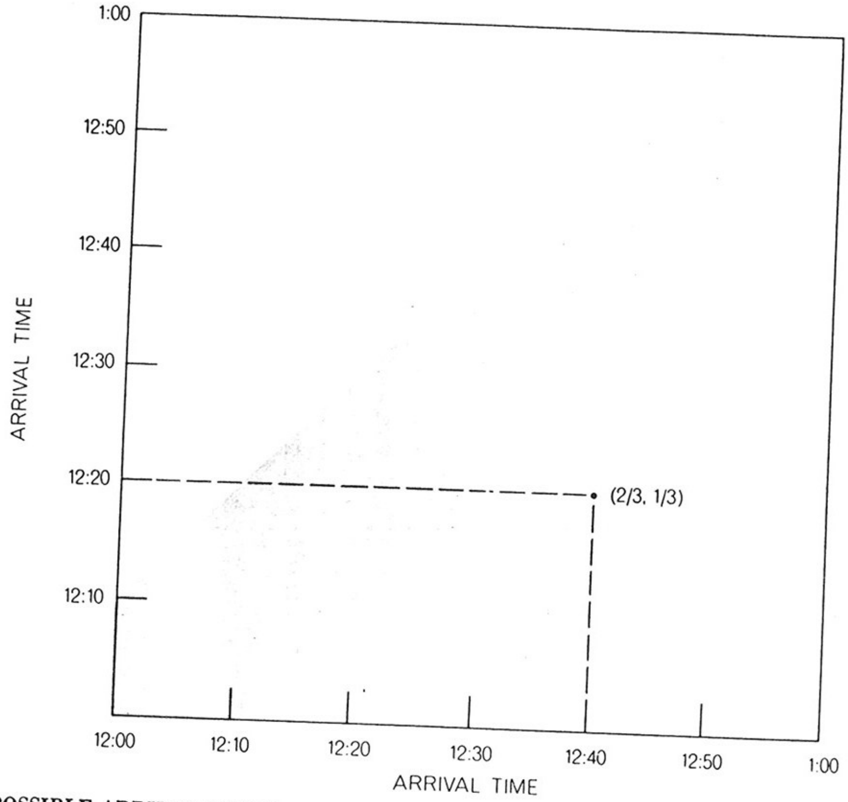
Why this apathy toward the subject among professional mathematicians? There were various reasons. The main one was the feeling that the entire theory seemed to be built on loose and nonrigorous foundations. Laplace's definition of probability, for instance, is based on the assumption that all the possible outcomes in question are equally likely; since this notion itself is a statement of probability, the definition appears to be a circular one. And that was not the worst objection. The field was plagued with apparent paradoxes and other difficulties. The rising standards of rigor in all branches of mathematics made probability seem an unprofitable subject to cultivate.

In the 1930's, however, it was restored to high standing among mathematicians by a significant clarification of its basic concepts and by its relation to measure theory, a branch of mathematics that goes back to Euclid and that early in this century was greatly extended and generalized by the French mathematicians Émile Borel and Henri Lebesgue. To understand and appreciate this development let us start with a celebrated problem in geometrical probability known as the Buffon needle problem. If a needle of a certain length (say four inches) is thrown at random on a floor made of planks wider than that length (say eight inches wide), what is the probability that the needle will fall across a crack between two planks? We can define the position of the needle at each throw by noting the location of the midpoint of the needle on a plank and the angle between the needle and a given crack [*see upper illustration at left*]. Now, we can also show the various possible positions of the needle by means of an abstract diagram in the form of a rectangle [*see lower illustration at left*], in which the height represents the width of the plank and the base represents the angle (in terms of π, with π equal to 180 degrees, $\pi/2$ equal to 90 degrees and so on).

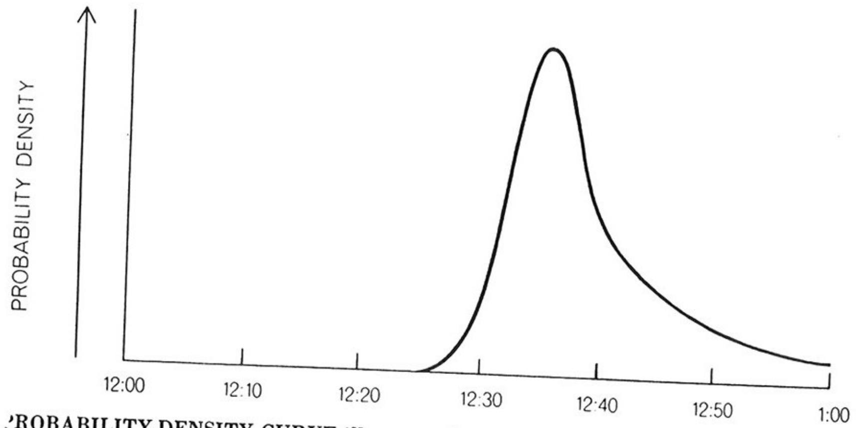
This rectangle as a whole, whose area is πd, represents all the possible positions in which the needle can fall. Technically it is called the "sample space," a general term used to denote all the possible outcomes in any probability experiment. (In tossing 10 coins the sample space is the set of all the 1,024 possible 10-item sequences of heads and/or tails.) In the needle experiment what part of the area of the rectangle corresponds to those positions of the needle in which it crosses a crack? This can be calculated by simple trigonometry, and it is represented by two sections within the rectangle with curved boundaries. Their combined area, which can be calculated by elementary calculus, turns out to be 2 times l, the length of the needle.

Now, if all the possible positions of the needle are equally likely, then the probability of the needle falling on a crack is the ratio of the dark colored areas in the illustration to the total area of the rectangle, or $2l/\pi d$. This is where the theory stumbles over its own arbitrary assumption. There is really no compelling reason to treat all the points in this abstract rectangle as equally likely, but the assumption is so natural as to appear inevitable. The degree of arbitrariness was dramatized by the French mathematician J. L. F. Bertrand, who devised examples (known as Bertrand paradoxes) in which, from assumptions that seemed equally natural, he could obtain quite different answers to a probability problem.

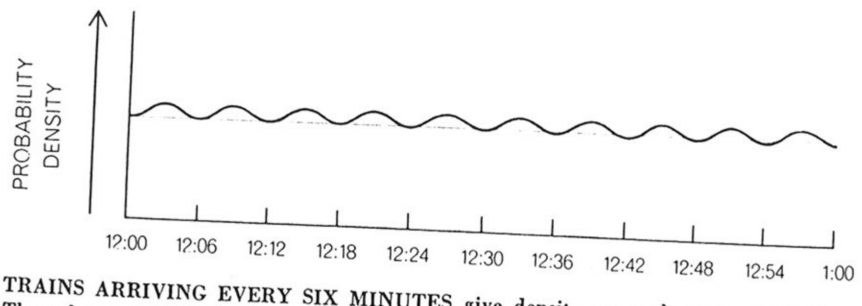
This was an unhappy situation, and a deeper understanding of the role and of the nature of probabilistic assumptions



POSSIBLE ARRIVAL TIMES of two suburbanites planning to meet at a library between 12:00 noon and 1:00 P.M. can be plotted. Arrival times for one person are on the vertical scale, for the other on the horizontal scale. Colored area covers region corresponding to a meeting. In order to meet they must arrive at library within 10 minutes of each other. As can be seen, if one arrives at 12:20 (a third of the way up) and the other at 12:40 (two-thirds of the way across), they will not meet. The (1/3, 2/3) point falls outside the shaded region.



PROBABILITY-DENSITY CURVE illustrates degree of unpredictability of arrival time if there is only one train, coming in at 12:20. Most likely meeting time at library is 12:35. Area of shaded portion represents probability of arrival between 12:40 and 12:44 P.M.



TRAINS ARRIVING EVERY SIX MINUTES give density curve shown by black line. The colored straight line is the "curve" in case all the arrival times are equally likely.

was called for. I can best explain the modern view of these matters by means of another example.

Suppose two friends living in different suburbs of New York City want to meet in front of the Forty-second Street Public Library at noontime. Railroad schedules (and performance) being what they are, the friends can only count on arriving sometime between 12:00 noon and 1:00 P.M. They agree to show up at the library somewhere in that interval, with the stipulation that, in order not to waste too much time waiting, each will wait only 10 minutes after arriving and then leave if the other has not shown up. What is the probability of their actually meeting?

I should mention that, although this case is admittedly artificial, it is by no means a trivial problem. Extended to many members instead of just two, it is analogous to (but far simpler than) an important unsolved problem in statistical mechanics whose solution would shed much light on the theory of changes of states of matter—for instance, from solid to liquid.

If we assume that each of the two friends may arrive anytime between 12:00 and 1:00, we can plot a geometrical "sample space" as in the Buffon needle problem. One person's possible times of arrival are denoted on the x axis, the other's on the y axis [*see top illustration on this page*]. We can then designate every possible pair of arrival times by a point in the square graph. Those points that lie within the part of the square that represents arrival times no more than 10 minutes apart will signify a meeting; all the other points will mean "no meeting." Taking the ratio of the two areas as the probability, as in the needle problem, we can calculate that the probability of the two friends meeting is 11/36—not quite one chance in three.

This case makes clear that we have made two different assumptions. Let us analyze them in a more general context.

In very general terms probability theory, as a mathematical discipline, is concerned with the problem of calculating the probabilities of complex events consisting of collections of "elementary" events whose probabilities are known or postulated. For example, in rolling two dice the appearance of a 10 is a "complex" event that consists of three elementary events: (1) the first die shows a 4 and the second a 6, (2) the first die shows a 5 and the second a 5 and

(3) the first die shows a 6 and the second a 4. The meeting of our two friends is also a complex event; an example of an elementary event would be the arrival of one of the friends in the interval between 12:20 and 12:25.

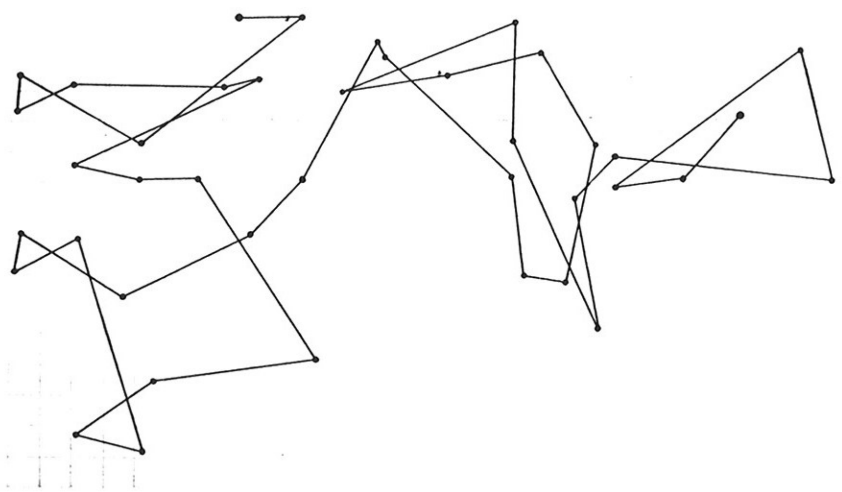
In our calculation of the probability of the two friends meeting, the first assumption we made was that each of the two individuals may arrive anytime between 12:00 and 1:00, all times of arrival being "equally likely." (The corresponding assumption about the dice is that any one of the six faces of each die may come up with equal probability.) But if each person is limited to only one train scheduled to arrive at Grand Central during the hour (say at 12:20 or later), this assumption is completely unrealistic: he will certainly not arrive in the early part of the hour. The situation corresponds to the two dice being loaded. On the other hand, if there are six scheduled trains, due to arrive at 10-minute intervals from 12:00 on, and if they tend to be haphazardly off schedule, the assumption becomes more reasonable, although it may still not be strictly correct to say that all the times of arrival are equally probable.

The second assumption we made was that the arrival times of the two friends are completely independent of each other. This assumption, like the first, is of crucial importance. In mathematical terms it is reflected in the rule of the multiplication of probabilities. This rule states that when individual events are independent of each other, the probability of the complex event that *all* of them will occur is the product of individual probabilities. (Actually from a strictly logical point of view the rule of the multiplication of probabilities constitutes a *definition* of independence.) Independence is assumed in the throw of two dice (which presumably are not linked in any way) as well as in the case of the two suburbanites coming into New York (provided that they have not "coupled" their arrival times by an understanding about the selection of particular trains).

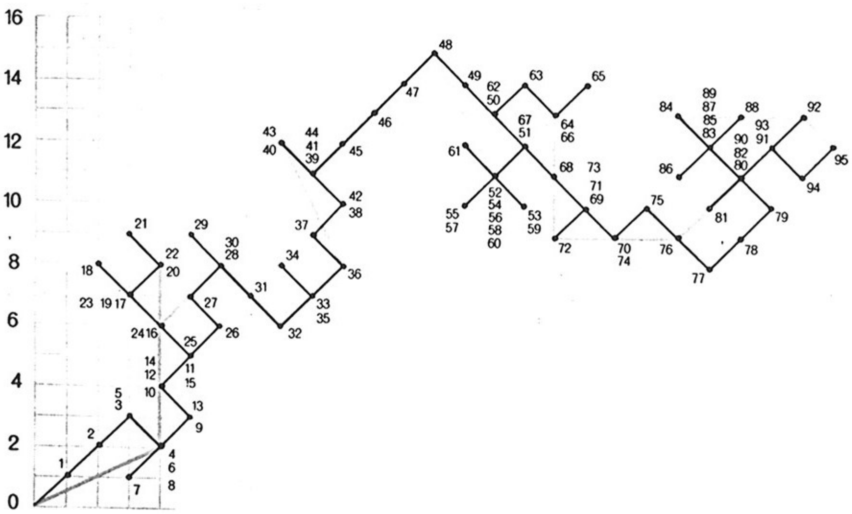
It should be noted that there is an important difference between the dice-throwing and suburbanite-meeting problems. In the first case the number of possible outcomes is finite (just 36), whereas in the second it is infinite, in the sense that the arrival times may occur at any instant within the hour; that is to say, the sample space is a continuum with an infinite number of "points."

To enable one to go on with calculations of probabilities two very general rules, or axioms, are introduced. The first concerns mutually exclusive events: events such that the occurrence of one precludes the occurrence of any other. For such events the probability that at least one will occur is the *sum* of individual probabilities (the axiom of additivity). The second concerns pairs of events such that one *implies* the other. In this case the probability that one will occur but not the other is obtained by subtracting the smaller probability from the larger.

Now, these rules for calculating probabilities of complex events are identical with those used for calculating areas and volumes in geometry. We can substitute the word "set" for "event" and "area" or "volume" for "probability." The problem then is to assign



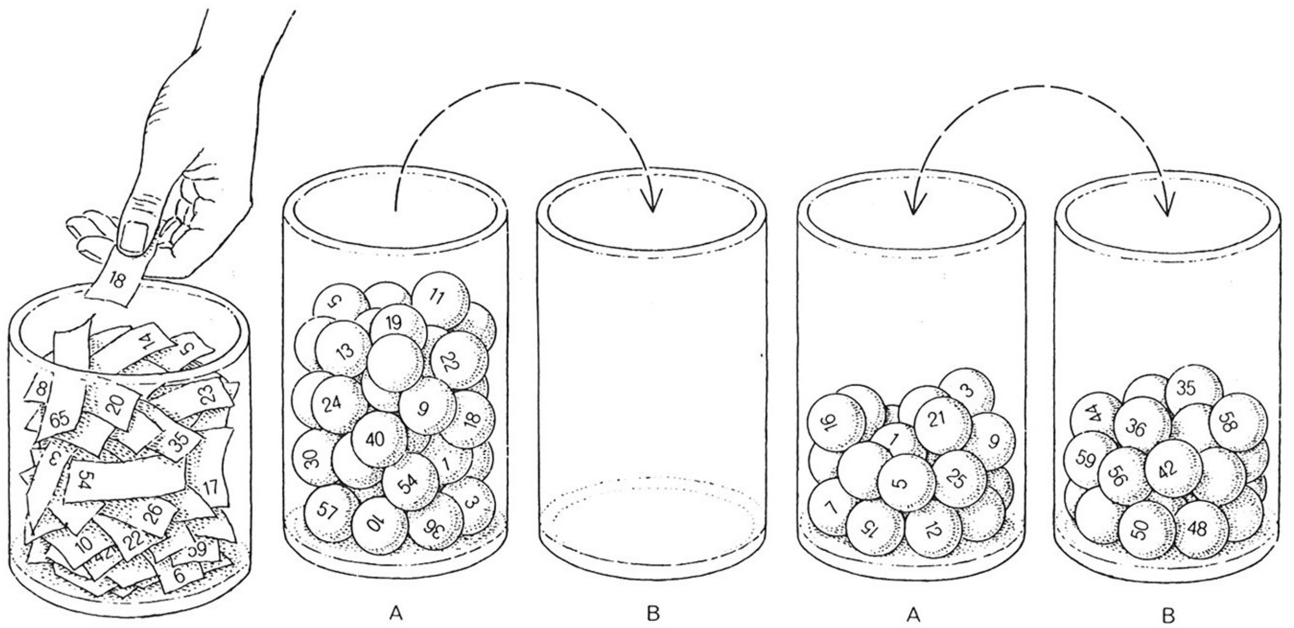
BROWNIAN PATH is taken by a particle being "kicked around" by molecules of a surrounding liquid or gas. A stochastic process (it varies continuously with time), Brownian motion can be analyzed and modeled (*illustration below*) by probabilistic techniques.



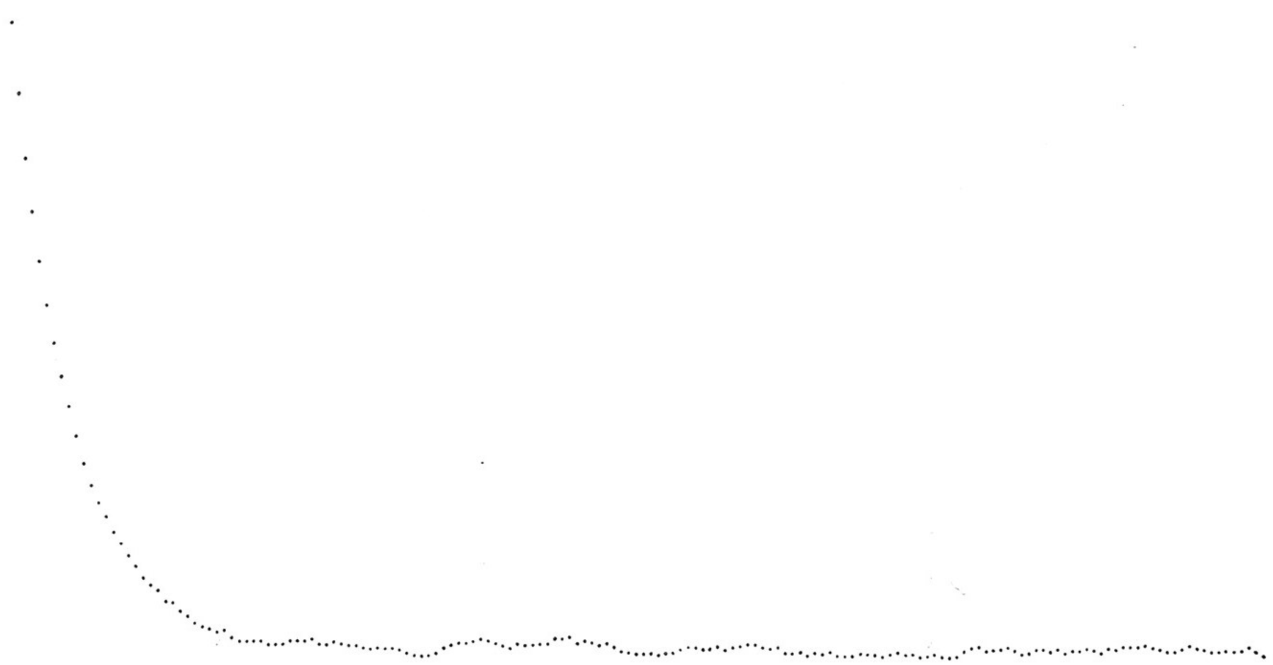
BROWNIAN MODEL can be constructed from records of coin tosses. Two series of 90-odd tosses of a coin were used, one plotted on the horizontal scale, the other on the vertical scale. The cumulative total of tails at each toss was subtracted from the cumulative total of heads. The first three tosses in both series were heads and the fourth toss in the horizontal series was heads, but the fourth toss in the vertical series was tails. Numbers on the dots are the toss numbers. Colored line traces "position" at every fourth toss, just as the track of a Brownian particle recorded by a camera shows only a small fraction of the staggering number of "kicks" such a particle receives from molecules around it.

appropriate areas to sets, and this is the province of measure theory, which has been given that name because the word "measure" is now used to refer to areas of very complex sets.

If we go back to the problem of the two suburbanite friends, we note that the set that corresponds to their meeting is quite simple. Its area, or probability, is well within Euclid's framework, and its calculation can be based on the manipulation of only a finite number of nonoverlapping rectangles. In the Buffon needle problem, since the region of interest is bounded by curves, one must allow an infinite number of rectangles, but the calculation of the area is still relatively simple and requires nothing more than elementary calculus. What was surprising and exciting about measure theory as it was de-



EHRENFEST MODEL for illustrating a Markov chain involves a game in which balls are moved from one container to another according to numbers drawn at random from a third container (*left*). As long as there are many more balls in container *A* than in container *B*, the flow of balls will be strongly from *A* to *B*. The probability of finding in *A* the ball with the drawn number changes in a way that depends on past drawings. This form of dependence of probability on past events is called a Markov chain.



PLAYED ON A COMPUTER, an Ehrenfest game with 16,384 hypothetical balls and 200,000 drawings took just two minutes. Starting with all the balls in container *A*, the number of balls in *A* was recorded with a dot every 1,000 drawings. It declined exponentially until equilibrium was reached with 8,192 balls (half of them) in each container. After that fluctuations were not great.

veloped by Borel and Lebesgue was that, by merely postulating that the measure of an *infinite* collection of unconnected sets should be the *sum* of the measures of the individual ones (corresponding to requiring that the probability that at least one out of *infinitely many* mutually exclusive events will occur should be the sum of individual probabilities), it was possible to assign measures to extremely complex sets.

Because of this, measure theory opened the way to the posing and solving of problems in probability that would have been unthinkable in Laplace's time. Here, for instance, is one of the problems that received much attention in the 1920's and 1930's and contributed greatly to bringing probability theory into the mainstream of mathematics.

Consider the infinite series

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}+\cdots$$

This is known as a diverging series; that is, by adding more and more terms one could exceed any given number.

Suppose the signs between the terms, instead of being all pluses, were made plus or minus at random by means of independent tosses of an honest coin. What is the probability that the resulting series would converge? That is to say, what is the probability that by extending the series to more and more terms one would come closer and closer to some terminating number?

To answer the question one must consider all the possible infinite sequences of heads and tails as the sample space. One sequence might begin: $H\,H\,T\,T\,T\,H\,T\,H\ldots$. If we let H represent plus and T represent minus, the number series above becomes

$$+\frac{1}{1}+\frac{1}{2}-\frac{1}{3}-\frac{1}{4}-\frac{1}{5}+\frac{1}{6}-\frac{1}{7}+\frac{1}{8}\cdots$$

With each such sequence we can associate a real number, t, between 0 and 1, and each t can be represented by a binary number in which the digit 1 denotes H and the digit 0 stands for T. The sequence cited above is then written as $t = .11000101\ldots$. The binary digits form a model of independent tosses of a coin. Now those t's that will yield convergent series form a set, and the probability that t falls into this set is the "measure" of the set. It turns out that the set of hypothetical t's that do *not* yield convergent series is so sparse that its measure, or probability, is zero (although the set has a very complex structure and is far from being empty). Hence the answer to the problem is that, when the series above is given random signs, the probability that it will converge is 1.

The foregoing is an example of problems in "denumerable probabilities," that is, those involving events described in discrete terms. During the past two decades mathematicians have pursued an even more productive investigation of the theory of "stochastic processes": the probabilistic analysis of phenomena that vary continuously in time. Stochastic processes arise in physics, astronomy, economics, genetics, ecology and many other fields of science. The simplest and most celebrated example of a stochastic process is the Brownian motion of a particle.

The late Norbert Wiener conceived the idea of basing the theory of Brownian motion on a theory of measure in a set of all continuous paths. This idea proved enormously fruitful for probability theory. It breathed new life into old problems such as that of determining the electrostatic potential of a conductor of "arbitrary" shape, a problem that occupied the minds of illustrious mathematicians for more than a century. More than that, it opened up entire new areas of research and led to fascinating connections between probability theory and other branches of mathematics.

A single article can only touch on a few of the main developments and sample problems in probability theory. The subject today embraces vast new fields such as information theory, the theory of queues, diffusion theory and mathematical statistics. One can sum up the position of probability in general by observing that it has become both an indispensable tool of the engineer and a thriving branch of pure mathematics now raised to a high level of formalism and rigor.

I want to close with a brief comment on the philosophical aspect of probability theory (in itself a vast subject on which many volumes have been written). The philosophical implications can be best illustrated by a specific case, and the one I shall discuss has to do with a conflict between the thermodynamic and the mechanical views of the behavior of matter.

Consider two containers, one containing gas, the other a vacuum. If the two containers are connected by a tube and a valve in the tube is suddenly opened, what happens? According to the second law of thermodynamics, gas rushes from container A into container B at an exponential rate until the pressure in the two containers is the same. This is an expression of the law of increasing entropy, which in its most pessimistic form predicts that ultimately all matter and energy in the universe will even out and settle down to what Rudolf Clausius, one of the fathers of the second law, called *Wärmetod* (heat death).

Now, the mechanical, or kinetic, view of matter pictures the situation in an entirely different way. True, the molecules of gas will tend to move from the region of higher pressure into the one of lower pressure, but the movement is not merely one-way. Bouncing against the walls and against one another, the molecules will take off in random directions, and those that travel into container B will be as likely to wander back to container A as to remain where they are. As a matter of fact, Poincaré showed in a mathematical theorem that a dynamical system such as this one would eventually return arbitrarily close to its original state, with all or virtually all the gas molecules back in container A.

In 1907 Paul and Tatiana Ehrenfest illustrated this idea with a simple and beautiful probabilistic model. Consider two containers, A and B, with a large number of numbered balls in A and none in B. From a container filled with numbered slips of paper pick a numeral at random (say 6) and then transfer the ball marked with that number from container A to container B. Put the slip of paper back and go on playing the game this way, each time drawing at random a number between 1 and N (the total number of balls originally in container A) and moving the ball of that number from the container where it happens to be to the other container [*see upper illustration on page 172*].

It is intuitively clear that as long as there are many more balls in A than there are in B the probability of drawing a number that corresponds to a ball in A will be considerably higher than vice versa. Thus the flow of balls at first will certainly be strongly from A to B. As the drawings continue, the probability of finding the drawn number in A will change in a way that depends on the past drawings. This form of dependence of probability on past events is called a Markov chain, and in the game we are considering, all pertinent facts can be explicitly and rigorously deduced. It turns out that, on an averaging basis, the number of balls in container A will indeed decrease at an exponential rate, as the thermodynamic theory predicts, until about half of the balls are in container B. But the calculation also shows that if the game is played long enough, then, with probability equal to 1, all the balls will eventually wind up back in container A, as Poincaré's theorem says!

How long, on the average, would it take to return to the initial state? The answer is 2^N drawings, which is a staggeringly large number even if the total number of balls (N) is as small as 100. This explains why behavior in nature, as we observe it, moves only in one direction instead of oscillating back and forth. The entire history of man is pitifully short compared with the time it would take for nature to reverse itself.

To test the theoretical calculations experimentally, the Ehrenfest game was played on a high-speed computer. It began with 16,384 "balls" in container A, and each run consisted of 200,000 drawings (which took less than two minutes on the computer). A curve was drawn showing the number of balls in A on the basis of the number recorded after every 1,000 drawings [*see lower illustration on page 172*]. As was to be expected, the curve of decline in the number of balls in A was almost perfectly exponential. After the number nearly reached the equilibrium level (that is, 8,192, or half the original number) the curve became wiggly, moving randomly up and down around that number. The wiggles were somewhat exaggerated by the vagaries of the machine itself, but they represented actual fluctuations that were bound to occur in the number of balls in A.

Those small, capricious fluctuations are models of the variability in nature and are all that stands between us and the heat death to which we are seemingly condemned by the second law of thermodynamics! Probability theory has reconciled the apparent conflict between the thermodynamic and the kinetic views of nature by showing that there is no real contradiction between them if the second law is interpreted flexibly. In fact, the development of the theory of probability in the 20th century has changed our attitudes to such an extent that we no longer expect the laws of nature to be construed rigidly or dogmatically.