CHAPTER III*

Fluctuations in Coin Tossing and Random Walks

This chapter digresses from our main topic, which is taken up again only in chapter V. Its material has traditionally served as a first orientation and guide to more advanced theories. Simple methods will soon lead us to results of far-reaching theoretical and practical importance. We shall encounter theoretical conclusions which not only are unexpected but actually come as a shock to intuition and common sense. They will reveal that commonly accepted notions concerning chance fluctuations are without foundation and that the implications of the law of large numbers are widely misconstrued. For example, in various applications it is assumed that observations on an individual coin-tossing game during a long time interval will yield the same statistical characteristics as the observation of the results of a huge number of independent games at one given instant. This is not so. Indeed, using a currently popular jargon we reach the conclusion that in a population of normal coins the majority is necessarily maladjusted. [For empirical illustrations see section 6 and example (4.*b*).]

Until recently the material of this chapter used to be treated by analytic methods and, consequently, the results appeared rather deep. The elementary method[1] used in the sequel is therefore a good example of the newly discovered power of combinatorial methods. The results are fairly representative of a wider class of fluctuation phenomena[2] to be discussed

* This chapter may be omitted or read in conjunction with the following chapters. Reference to its contents will be made in chapters X (laws of large numbers), XI (first-passage times), XIII (recurrent events), and XIV (random walks), but the contents will not be used explicitly in the sequel.

[1] The discovery of the possibility of an elementary approach was the principal motivation for the second edition of this book (1957). The present version is new and greatly improved since it avoids various combinatorial tricks.

[2] See footnote 12.

in volume 2. All results will be derived anew, independently, by different methods. This chapter will therefore serve primarily readers who are not in a hurry to proceed with the systematic theory, or readers interested in the spirit of probability theory without wanting to specialize in it. For other readers a comparison of methods should prove instructive and interesting. Accordingly, *the present chapter should be read at the reader's discretion independently of, or parallel to, the remainder of the book.*

1. GENERAL ORIENTATION. THE REFLECTION PRINCIPLE

From a formal point of view we shall be concerned with arrangements of finitely many plus ones and minus ones. Consider $n = p + q$ symbols $\epsilon_1, \ldots, \epsilon_n$, each standing either for $+1$ or for -1; suppose that there are p plus ones and q minus ones. The partial sum $s_k = \epsilon_1 + \cdots + \epsilon_k$ represents the difference between the number of pluses and minuses occurring at the first k places. Then

$$(1.1) \qquad s_k - s_{k-1} = \epsilon_k = \pm 1, \qquad s_0 = 0, \qquad s_n = p - q,$$

where $k = 1, 2, \ldots, n$.

We shall use a geometric terminology and refer to rectangular coordinates t, x; for definiteness we imagine the t-axis is horizontal, the x-axis vertical. The arrangement $(\epsilon_1, \ldots, \epsilon_n)$ will be represented by a polygonal line whose kth side has slope ϵ_k and whose kth vertex has ordinate s_k. Such lines will be called paths.

Definition. *Let* $n > 0$ *and* x *be integers. A path* $(s_1, s_2, \ldots, s_n)$ *from the origin to the point* (n, x) *is a polygonal line whose vertices have abscissas* $0, 1, \ldots, n$ *and ordinates* $s_0, s_1, \ldots, s_n$ *satisfying* (1.1) *with* $s_n = x$.

We shall refer to n as the *length* of the path. There are 2^n paths of length n. If p among the ϵ_k are positive and q are negative, then

$$(1.2) \qquad n = p + q, \qquad x = p - q.$$

A path from the origin to an arbitrary point (n, x) exists only if n and x are of the form (1.2). In this case the p places for the positive ϵ_k can be chosen from the $n = p + q$ available places in

$$(1.3) \qquad N_{n,x} = \binom{p+q}{p} = \binom{p+q}{q}$$

different ways. For convenience we *define* $N_{n,x} = 0$ *whenever* n *and* x

are not of the form (1.2). With this convention *there exist exactly* $N_{n,x}$ *different paths from the origin to an arbitrary point* (n, x).

Before turning to the principal topic of this chapter, namely the theory of random walks, we illustrate possible applications of our scheme.

Examples. (*a*) *The ballot theorem.* The following amusing proposition was proved in 1878 by W. A. Whitworth, and again in 1887 by J. Bertrand.

Suppose that, in a ballot, candidate P *scores* p *votes and candidate* Q *scores* q *votes, where* $p > q$. *The probability that throughout the counting there are always more votes for* P *than for* Q *equals* $(p-q)/(p+q)$.

Similar problems of arrangements have attracted the attention of students of combinatorial analysis under the name of ballot problems. The recent renaissance of combinatorial methods has increased their popularity, and it is now realized that a great many important problems may be reformulated as variants of some generalized ballot problem.[3]

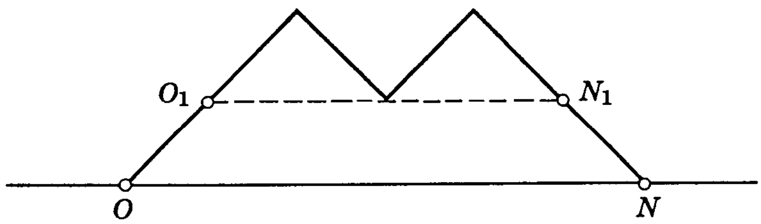

Figure 1. Illustrating positive paths. The figure shows also that there are exactly as many strictly positive paths from the origin to the point $(2n, 0)$ as there are non-negative paths from the origin to $(2n-2, 0)$.

The whole voting record may be represented by a path of length $p + q$ in which $\epsilon_k = +1$ if the kth vote is for P; conversely, every path from the origin to the point $(p + q, p - q)$ can be interpreted as a record of a voting with the given totals p and q. Clearly s_k is the number of votes by which P leads, or trails, just after the kth vote is cast. The candidate P leads throughout the voting if, and only if, $s_1 > 0, \ldots, s_n > 0$, that is, if all vertices lie strictly above the t-axis. (The path from 0 to N_1 in figure 1 is of this type.) The ballot theorem assumes tacitly that all admissible paths are equally probable. The assertion then reduces to the theorem proved at the end of this section as an immediate consequence of the reflection lemma.

(*b*) *Galton's rank order test.*[4] Suppose that a quantity (such as the height

[3] A survey of the history and the literature may be found in *Some aspects of the random sequence*, by D. E. Barton and C. L. Mallows [Ann. Math. Statist., vol. 36 (1965), pp. 236–260]. These authors discuss also various applications. The most recent generalization with many applications in queuing theory is due to L. Takacs.

[4] J. L. Hodges, Biometrika, vol. 42 (1955), pp. 261–262.

of plants) is measured on each of r treated subjects, and also on each of r control subjects. Denote the measurements by $a_1, \ldots, a_r$ and $b_1, \ldots, b_r$, respectively. To fix ideas, suppose that each group is arranged in decreasing order: $a_1 > a_2 > \cdots$ and $b_1 > b_2 > \ldots$. (To avoid trivialities we assume that no two observations are equal.) Let us now combine the two sequences into one sequence of $n = 2r$ numbers arranged in decreasing order. For an extremely successful treatment all the a's should precede the b's, whereas a completely ineffectual treatment should result in a random placement of a's and b's. Thus the efficiency of the treatment can be judged by the number of different a's that precede the b of the same rank, that is, by the number of subscripts k for which $a_k > b_k$. This idea was first used in 1876 by F. Galton for data referred to him by Charles Darwin. In this case r equaled 15 and the a's were ahead 13 times. Without knowledge of the actual probabilities Galton concluded that the treatment *was* effective. But, assuming perfect randomness, the probability that the a's lead 13 times or more equals $\frac{3}{16}$. This means that in three out of sixteen cases a perfectly ineffectual treatment would appear as good or better than the treatment classified as effective by Galton. This shows that a quantitative analysis may be a valuable supplement to our rather shaky intuition.

For an interpretation in terms of paths write $\epsilon_k = +1$ or -1 according as the kth term of the combined sequence is an a or a b. The resulting path of length $2r$ joins the origin to the point $(2r, 0)$ of the t-axis. The event $a_k > b_k$ occurs if, and only if, s_{2k-1} contains at least k plus ones, that is, if $s_{2k-1} > 0$. This entails $s_{2k} \geq 0$, and so the $(2k-1)$st and the $2k$th sides are above the t-axis. It follows that the inequality $a_k > b_k$ holds ν times if, and only if, 2ν sides lie above the t-axis. In section 9 we shall prove the unexpected result that the probability for this is $1/(r+1)$, irrespective of ν. (For related tests based on the theory of runs see II, 5.*b*.)

(*c*) *Tests of the Kolmogorov-Smirnov type.* Suppose that we observe two populations of the same biological species (animals or plants) living at different places, or that we wish to compare the outputs of two similar machines. For definiteness let us consider just one measurable characteristic such as height, weight, or thickness, and suppose that for each of the two populations we are given a sample of r observations, say $a_1, \ldots, a_r$ and $b_1, \ldots, b_r$. The question is roughly whether these data are consistent with the hypothesis that the two populations are statistically identical. In this form the problem is vague, but for our purposes it is not necessary to discuss its more precise formulation in modern statistical theory. It suffices to say that the tests are based on a comparison of the two empirical distributions. For every t denote by $A(t)$ the fraction k/n of subscripts i for which $a_i \leq t$. The function so defined over the real axis is the *empirical distribution* of the a's. The empirical distribution B is defined in like manner. A refined mathematical theory originated by N. V. Smirnov (1939) derives the probability distribution of the maximum of the discrepancies $|A(t) - B(t)|$ and of other quantities which can be used for testing the stated hypothesis. The theory is rather intricate, but was greatly simplified and made more intuitive by B. V. Gnedenko who had the lucky idea to connect it with the geometric theory of paths. As in the preceding example we associate with the two samples a path of length $2r$ leading from the origin to the point $(2r, 0)$. To say that the two populations are statistically indistinguishable amounts to saying that ideally the sampling experiment makes all possible paths equally probable. Now it is easily seen that $|A(t) - B(t)| > \xi$ for some t if, and only if, $|s_k| > \xi r$ for some k. The probability of this event is simply the probability that a path of length $2r$ leading from the origin to the point $(0, 2r)$ is not constrained to the interval between $\pm\xi r$. This probability has been known for a long time because it is connected with the ruin problem in random walks and with the physical problem of diffusion with absorbing barriers. (See problem 3.)

This example is beyond the scope of the present volume, but it illustrates how random walks can be applied to problems of an entirely different nature.

(*d*) *The ideal coin-tossing game and its relation to stochastic processes.* A path of length n can be interpreted as the record of an ideal experiment consisting of n successive tosses of a coin. If $+1$ stands for heads, then s_k equals the (positive or negative) excess of the accumulated number of heads over tails at the conclusion of the kth trial. The classical description introduces the fictitious gambler Peter who at each trial wins or loses a unit amount. The sequence $s_1, s_2, \ldots, s_n$ then represents Peter's successive cumulative gains. It will be seen presently that they are subject to chance fluctuations of a totally unexpected character.

The picturesque language of gambling should not detract from the general importance of the coin-tossing model. In fact, the model may serve as a first approximation to many more complicated chance-dependent processes in physics, economics, and learning theory. Quantities such as the energy of a physical particle, the wealth of an individual, or the accumulated learning of a rat are supposed to vary in consequence of successive collisions or random impulses of some sort. For purposes of a first orientation one assumes that the individual changes are of the same magnitude, and that their sign is regulated by a coin-tossing game. Refined models take into account that the changes and their probabilities vary from trial to trial, but even the simple coin-tossing model leads to surprising, indeed to shocking, results. They are of practical importance because they show that, contrary to generally accepted views, the laws governing a prolonged series of individual observations will show patterns and averages far removed from those derived for a whole population. In other words, currently popular psychological tests would lead one to say that in a population of "normal" coins most individual coins are "maladjusted."

It turns out that the chance fluctuations in coin tossing are typical for more general chance processes with cumulative effects. Anyhow, it stands to reason that if even the simple coin-tossing game leads to paradoxical results that contradict our intuition, the latter cannot serve as a reliable guide in more complicated situations. ◀

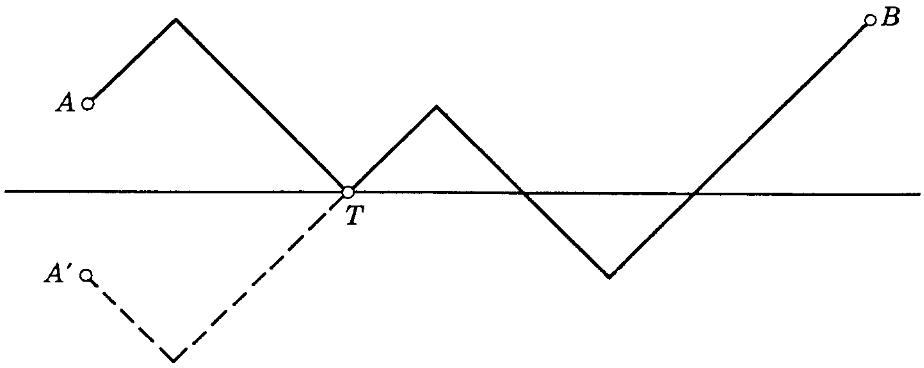

Figure 2. Illustrating the reflection principle.

It is as surprising as it is pleasing that most important conclusions can be drawn from the following simple lemma.

Let $A = (a, \alpha)$ and $B = (b, \beta)$ be integral points in the positive quadrant: $b > a \geq 0$, $\alpha > 0$, $\beta > 0$. By reflection of A on the t-axis is meant the point $A' = (a, -\alpha)$. (See figure 2.) A path from A to B is defined in the obvious manner.

Lemma.[5] (*Reflection principle.*) *The number of paths from A to B which touch or cross the x-axis equals the number of all paths from A' to B.*

Proof. Consider a path $(s_a = \alpha, s_{a+1}, \ldots, s_b = \beta)$ from A to B having one or more vertices on the t-axis. Let t be the abscissa of the first such vertex (see figure 2); that is, choose t so that $s_a > 0, \ldots, s_{t-1} > 0$, $s_t = 0$. Then $(-s_a, -s_{a+1}, \ldots, -s_{t-1}, s_t = 0, s_{t+1}, s_{t+2}, \ldots, s_b)$ is a

[5] The reflection principle is used frequently in various disguises, but without the geometrical interpretation it appears as an ingenious but incomprehensible trick. The probabilistic literature attributes it to D. André (1887). It appears in connection with the difference equations for random walks in XIV, 9. These are related to some partial differential equations where the reflection principle is a familiar tool called *method of images*. It is generally attributed to Maxwell and Lord Kelvin. For the use of repeated reflections see problems 2 and 3.

path leading from A' to B and having $T = (t, 0)$ as its first vertex on the t-axis. The sections AT and $A'T$ being reflections of each other, there exists a one-to-one correspondence between all paths from A' to B and such paths from A to B that have a vertex on the x-axis. This proves the lemma. ▶

As an immediate consequence we prove the result discussed in example (*a*). It will serve as starting point for the whole theory of this chapter.

The ballot theorem. *Let n and x be positive integers. There are exactly $\frac{x}{n} N_{n,x}$ paths $(s_1, \ldots, s_n = x)$ from the origin to the point (n, x) such that $s_1 > 0, \ldots, s_n > 0$.*

Proof. Clearly there exist exactly as many admissible paths as there are paths from the point $(1, 1)$ to (n, x) which neither touch or cross the t-axis. By the last lemma the number of such paths equals

$$N_{n-1,x-1} - N_{n-1,x+1} = \binom{p+q-1}{p-1} - \binom{p+q-1}{p}$$

with p and q defined in (1.2). A trite calculation shows that the right side equals $N_{n,x}(p-q)/(p+q)$, as asserted. ▶

2. RANDOM WALKS: BASIC NOTIONS AND NOTATIONS

The ideal coin-tossing game will now be described in the terminology of random walks which has greater intuitive appeal and is better suited for generalizations. As explained in the preceding example, when a path $(s_1, \ldots, s_\rho)$ is taken as record of ρ successive coin tossings the partial sums $s_1, \ldots, s_\rho$ represent the successive cumulative gains. For the geometric description it is convenient to pretend that the tossings are performed at a uniform rate so that the nth trial occurs at epoch[6] n. The successive partial sums $s_1, \ldots, s_n$ will be marked as points on the vertical x-axis; they will be called the positions of a "particle" performing a random walk. Note that the particle moves in unit steps, up or down, on a

[6] Following J. Riordan, the word *epoch* is used to denote *points* on the time axis because some contexts use the alternative terms (such as moment, time, point) in different meanings. Whenever used mathematically, the word time will refer to an interval or duration. A physical experiment may take some time, but our ideal trials are timeless and occur at epochs.

line. A path represents the record of such a movement. For example, the path from O to N in figure 1 stands for a random walk of six steps terminating by a return to the origin.

Each path of length ρ can be interpreted as the outcome of a random walk experiment; there are 2^ρ such paths, and we attribute probability $2^{-\rho}$ to each. (Different assignments will be introduced in chapter XIV. To distinguish it from others the present random walk is called *symmetric.*)

We have now completed the definition of the sample space and of the probabilities in it, but the dependence on the unspecified number ρ is disturbing. To see its role consider the event that the path passes through the point $(2, 2)$. The first two steps must be positive, and there are $2^{\rho-2}$ paths with this property. As could be expected, the probability of our event therefore equals $\frac{1}{4}$ regardless of the value of ρ. More generally, for any $k \leq \rho$ it is possible to prescribe arbitrarily the first k steps, and exactly $2^{\rho-k}$ paths will satisfy these k conditions. It follows that *an event determined by the first $k \leq \rho$ steps has a probability independent of* ρ. In practice, therefore, the number ρ plays no role provided it is sufficiently large. In other words, any path of length n can be taken as the initial section of a very long path, and there is no need to specify the latter length. Conceptually and formally it is most satisfactory to consider unending sequences of trials, but this would require the use of non-denumerable sample spaces. In the sequel it is therefore understood that the length ρ of the paths constituting the sample space is larger than the number of steps occurring in our formulas. Except for this we shall be permitted, and glad, to forget about ρ.

To conform with the notations to be used later on in the general theory we shall denote the individual steps generically by $\mathbf{X}_1, \mathbf{X}_2, \ldots$ and the positions of the particle by $\mathbf{S}_1, \mathbf{S}_2, \ldots$. Thus

$$\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n, \qquad \mathbf{S}_0 = 0. \tag{2.1}$$

From any particular path one can read off the corresponding values of $\mathbf{X}_1, \mathbf{X}_2, \ldots$; that is, the $\mathbf{X}_k$ are functions of the path.[7] For example, for the path of figure 1 clearly $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}_4 = 1$ and $\mathbf{X}_3 = \mathbf{X}_5 = \mathbf{X}_6 = -1$.

We shall generally describe all events by stating the appropriate conditions on the sums $\mathbf{S}_k$. Thus the event "at epoch n the particle is at the point r" will be denoted by $\{\mathbf{S}_n = r\}$. For its probability we write $p_{n,r}$. (For smoother language we shall describe this event as a "visit" to r at

[7] In the terminology to be introduced in chapter IX the $\mathbf{X}_k$ are random variables.

epoch n.) The number $N_{n,r}$ of paths from the origin to the point (n, r) is given by (1.3), and hence

$$p_{n,r} = \mathbf{P}\{\mathbf{S}_n = r\} = \binom{n}{\frac{n+r}{2}} 2^{-n}, \tag{2.2}$$

where it is understood that the binomial coefficient is to be interpreted as zero unless $(n+r)/2$ is an integer between 0 and n, inclusive.

A *return to the origin* occurs at epoch k if $\mathbf{S}_k = 0$. Here k is necessarily even, and for $k = 2\nu$ the probability of a return to the origin equals $p_{2\nu,0}$. Because of the frequent occurrence of this probability we denote it by $u_{2\nu}$. Thus

$$u_{2\nu} = \binom{2\nu}{\nu} 2^{-2\nu}. \tag{2.3}$$

When the binomial coefficient is expressed in terms of factorials, Stirling's formula II, (9.1) shows directly that

$$u_{2\nu} \sim \frac{1}{\sqrt{\pi\nu}} \tag{2.4}$$

where the sign $\sim$ indicates that the ratio of the two sides tends to 1 as $\nu \to \infty$; the right side serves as excellent approximation[8] to $u_{2\nu}$ even for moderate values of ν.

Among the returns to the origin the *first return* commands special attention. A first return occurs at epoch 2ν if

$$\mathbf{S}_1 \neq 0, \ldots, \mathbf{S}_{2\nu-1} \neq 0, \quad \textit{but} \quad \mathbf{S}_{2\nu} = 0. \tag{2.5}$$

The probability for this event will be denoted by $f_{2\nu}$. By definition $f_0 = 0$.

The probabilities f_{2n} and u_{2n} are related in a noteworthy manner. A visit to the origin at epoch $2n$ may be the first return, or else the first return occurs at an epoch $2k < 2n$ and is followed by a renewed return $2n - 2k$ time units later. The probability of the latter contingency is $f_{2k}u_{2n-2k}$ because there are $2^{2k}f_{2k}$ paths of length $2k$ ending with a first return, and $2^{2n-2k}u_{2n-2k}$ paths from the point $(2k, 0)$ to $(2n, 0)$. It follows that

$$u_{2n} = f_2u_{2n-2} + f_4u_{2n-4} + \cdots + f_{2n}u_0, \qquad n \geq 1. \tag{2.6}$$

(See problem 5.)

[8] For the true value $u_{10} = 0.2461$ we get the approximation 0.2523; for $u_{20} = 0.1762$ the approximation is 0.1784. The per cent error decreases roughly in inverse proportion to ν.

The normal approximation. Formula (2.2) gives no direct clue as to the range within which S_n is likely to fall. An answer to this question is furnished by an approximation formula which represents a special case of the central limit theorem and will be proved[9] in VII, 2.

The probability that $a < S_n < b$ is obtained by summing probabilities $p_{n,r}$ over all r between a and b. For the evaluation it suffices to know the probabilities for all inequalities of the form $S_n > a$. Such probabilities can be estimated from the fact that for all x as $n \to \infty$

$$\mathbf{P}\{S_n > x\sqrt{n}\} \to 1 - \mathfrak{N}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{1}{2}t^2}\, dt \tag{2.7}$$

where $\mathfrak{N}$ stands for the normal distribution function defined in VII, 1. Its nature is of no particular interest for our present purposes. The circumstance that the limit exists shows the important fact that for large n the ratios $S_n/\sqrt{n}$ are governed approximately by the same probabilities and so the same approximation can be used for all large n.

The accompanying table gives a good idea of the probable range of S_n. More and better values will be found in table 1 of chapter VII.

TABLE 1

x	0.5	1.0	1.5	2.0	2.5	3.0
$\mathbf{P}\{S_n > x\sqrt{n}\}$	0.309	0.159	0.067	0.023	0.006	0.001

3. THE MAIN LEMMA

As we saw, the probability of a return to the origin at epoch 2ν equals the quantity $u_{2\nu}$ of (2.3). As the theory of fluctuations in random walks began to take shape it came as a surprise that almost all formulas involved this probability. One reason for this is furnished by the following simple lemma, which has a mild surprise value of its own and provides the key to the deeper theorems of the next section.

Lemma 1.[10] *The probability that no return to the origin occurs up to and including epoch $2n$ is the same as the probability that a return occurs at epoch $2n$. In symbols,*

$$\mathbf{P}\{S_1 \neq 0, \ldots, S_{2n} \neq 0\} = \mathbf{P}\{S_{2n} = 0\} = u_{2n}. \tag{3.1}$$

[9] The special case required in the sequel is treated *separately* in VII, 2 without reference to the general binomial distribution. The proof is simple and can be inserted at this place.

[10] This lemma is obvious from the form of the generating function $\Sigma f_{2k}s^{2k}$ [see XI, (3.6)] and has been noted for its curiosity value. The discovery of its significance is recent. For a geometric proof see problem 7.

Here, of course, $n > 0$. When the event on the left occurs either all the $\mathbf{S}_j$ are positive, or all are negative. The two contingencies being equally probable we can restate (3.1) in the form

$$\mathbf{P}\{\mathbf{S}_1 > 0, \ldots, \mathbf{S}_{2n} > 0\} = \tfrac{1}{2}u_{2n}. \tag{3.2}$$

Proof. Considering all the possible values of $\mathbf{S}_{2n}$ it is clear that

$$\mathbf{P}\{\mathbf{S}_1 > 0, \ldots, \mathbf{S}_{2n} > 0\} = \sum_{r=1}^{\infty} \mathbf{P}\{\mathbf{S}_1 > 0, \ldots, \mathbf{S}_{2n-1} > 0, \mathbf{S}_{2n} = 2r\} \tag{3.3}$$

(where all terms with $r > n$ vanish). By the ballot theorem the number of paths satisfying the condition indicated on the right side equals $N_{2n-1,2r-1} - N_{2n-1,2r+1}$, and so the rth term of the sum equals

$$\tfrac{1}{2}(p_{2n-1,2r-1} - p_{2n-1,2r+1}).$$

The negative part of the rth term cancels against the positive part of the $(r+1)$st term with the result that the sum in (3.3) reduces to $\tfrac{1}{2}p_{2n-1,1}$. It is easily verified that $p_{2n-1,1} = u_{2n}$ and this concludes the proof. ►

The lemma can be restated in several ways; for example,

$$\mathbf{P}\{\mathbf{S}_1 \geq 0, \ldots, \mathbf{S}_{2n} \geq 0\} = u_{2n}. \tag{3.4}$$

Indeed, a path of length $2n$ with all vertices strictly above the x-axis passes through the point $(1, 1)$. Taking this point as new origin we obtain a path of length $2n - 1$ with all vertices above or on the new x-axis. It follows that

$$\mathbf{P}\{\mathbf{S}_1 > 0, \ldots, \mathbf{S}_{2n} > 0\} = \tfrac{1}{2}\mathbf{P}\{\mathbf{S}_1 \geq 0, \ldots, \mathbf{S}_{2n-1} \geq 0\}. \tag{3.5}$$

But $\mathbf{S}_{2n-1}$ is an odd number, and hence $\mathbf{S}_{2n-1} \geq 0$ implies that also $\mathbf{S}_{2n} \geq 0$. The probability on the right in (3.5) is therefore the same as (3.4) and hence (3.4) is true. (See problem 8.)

Lemma 1 leads directly to an explicit expression for the probability distribution for the first return to the origin. Saying that a first return occurs at epoch $2n$ amounts to saying that the conditions

$$\mathbf{S}_1 \neq 0, \ldots, \mathbf{S}_{2k} \neq 0$$

are satisfied for $k = n - 1$, but not for $k = n$. In view of (3.1) this means that

$$f_{2n} = u_{2n-2} - u_{2n}, \qquad n = 1, 2, \ldots. \tag{3.6}$$

A trite calculation reduces this expression to

$$f_{2n} = \frac{1}{2n-1}\, u_{2n}. \tag{3.7}$$

We have thus proved

Lemma 2. *The probability that the first return to the origin occurs at epoch* $2n$ *is given by* (3.6) *or* (3.7).

It follows from (3.6) that $f_2 + f_4 + \cdots = 1$. In the coin-tossing terminology this means that an ultimate equalization of the fortunes becomes practically certain if the game is prolonged sufficiently long. This was to be anticipated on intuitive grounds, except that the great number of trials necessary to achieve practical certainty comes as a surprise. For example, the probability that no equalization occurs in 100 tosses is about 0.08.

4. LAST VISIT AND LONG LEADS

We are now prepared for a closer analysis of the nature of chance fluctuations in random walks. The results are startling. According to widespread beliefs a so-called law of averages should ensure that in a long coin-tossing game each player will be on the winning side for about half the time, and that the lead will pass not infrequently from one player to the other. Imagine then a huge sample of records of ideal coin-tossing games, each consisting of exactly $2n$ trials. We pick one at random and observe the epoch of the last tie (in other words, the number of the last trial at which the accumulated numbers of heads and tails were equal). This number is even, and we denote it by $2k$ (so that $0 \leq k \leq n$). Frequent changes of the lead would imply that k is likely to be relatively close to n, but this is not so. Indeed, the next theorem reveals the amazing fact that the distribution of k is symmetric in the sense that any value k has exactly the same probability as $n - k$. This symmetry implies in particular that the inequalities $k > n/2$ and $k < n/2$ are equally likely.[11] *With probability* $\frac{1}{2}$ *no equalization occurred in the second half of the game, regardless of the length of the game.* Furthermore, the probabilities near the end points are *greatest*; the most probable values for k are the extremes 0 and n. These results show that intuition leads to an erroneous picture of the probable effects of chance fluctuations. A few numerical results may be illuminating.

---

[11] The symmetry of the distribution for k was found empirically by computers and verified theoretically without knowledge of the exact distribution (4.1). See D. Blackwell, P. Dewel, and D. Freedman, Ann. Math. Statist., vol. 35 (1964), p. 1344.

Examples. (*a*) Suppose that a great many coin-tossing games are conducted simultaneously at the rate of one per second, day and night, for a whole year. On the average, in one out of ten games the last equalization will occur before 9 days have passed, and the lead will not change during the following 356 days. In one out of twenty cases the last equalization takes place within $2\frac{1}{4}$ days, and in one out of a hundred cases it occurs within the first 2 hours and 10 minutes.

(*b*) Suppose that in a learning experiment lasting one year a child was consistently lagging except, perhaps, during the initial week. Another child was consistently ahead except, perhaps, during the last week. Would the two children be judged equal? Yet, let a group of 11 children be exposed to a similar learning experiment involving no intelligence but only chance. One among the 11 would appear as leader for all but one week, another as laggard for all but one week.

The exact probabilities for the possible values of k are given by

Theorem 1. (*Arc sine law for last visits.*) *The probability that up to and including epoch* $2n$ *the last visit to the origin occurs at epoch* $2k$ *is given by*

$$\alpha_{2k,2n} = u_{2k}u_{2n-2k}, \qquad k = 0, 1, \ldots, n. \tag{4.1}$$

Proof. We are concerned with paths satisfying the conditions $S_{2k} = 0$ and $S_{2k+1} \neq 0, \ldots, S_{2n} \neq 0$. The first $2k$ vertices can be chosen in $2^{2k}u_{2k}$ different ways. Taking the point $(2k, 0)$ as new origin and using (3.1) we see that the next $(2n-2k)$ vertices can be chosen in $2^{2n-2k}u_{2n-2k}$ ways. Dividing by 2^{2n} we get (4.1). ▶

It follows from the theorem that the numbers (4.1) add to unity. The probability distribution which attaches weight $\alpha_{2k,2n}$ to the point $2k$ will be called *the discrete arc sine distribution of order* n, because the inverse sine function provides excellent numerical approximations. The distribution is symmetric in the sense that $\alpha_{2k,2n} = \alpha_{2n-2k,2n}$. For $n = 2$ the three values are $\frac{3}{8}$, $\frac{2}{8}$, $\frac{3}{8}$; for $n = 10$ see table 2. The central term is always smallest.

The main features of the arc sine distributions are best explained by

TABLE 2

DISCRETE ARC SINE DISTRIBUTION OF ORDER 10

	$k = 0$ $k = 10$	$k = 1$ $k = 9$	$k = 2$ $k = 8$	$k = 3$ $k = 7$	$k = 4$ $k = 6$	$k = 5$
$\alpha_{2k,20}$	0.1762	0.0927	0.0736	0.0655	0.0617	0.0606

means of the graph of the function

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}} \qquad 0 < x < 1. \tag{4.2}$$

Using Stirling's formula it is seen that u_{2n} is close to $1/\sqrt{\pi n}$, except when

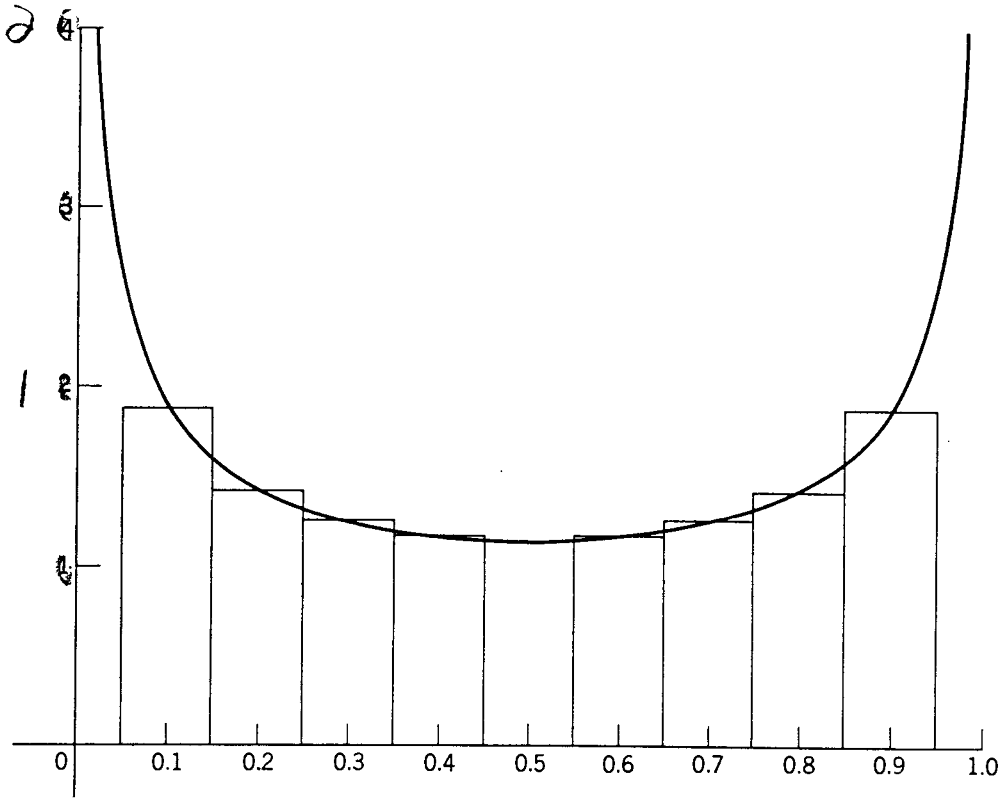

Figure 3. Graph of $f(x) = \dfrac{1}{\pi\sqrt{x(1-x)}}$. The construction explains the approximation (4.3).

n is very small. This yields the approximation

$$\alpha_{2k,2n} \approx \frac{1}{n} f(x_k), \qquad \textit{where} \quad x_k = \frac{k}{n}; \tag{4.3}$$

the error committed is negligible except when k is extremely close to 0 or n. The right side equals the area of a rectangle with height $f(x_k)$ whose basis is the interval of length $1/n$ centered at x_k (see figure 3). For $0 < p < q < 1$ and large n the sum of the probabilities $\alpha_{2k,2n}$ with $pn < k < qn$ is therefore approximately equal to the area under the graph of f and above the interval $p < x < q$. This remains true also for $p = 0$ and $q = 1$ because the total area under the graph equals unity which is also true of the sum over all $\alpha_{2k,2n}$. Fortunately (4.2) can be integrated explicitly and we conclude that *for fixed* $0 < x < 1$ *and* n *sufficiently large*

$$\sum_{k < xn} \alpha_{2k,2n} \approx \frac{2}{\pi} \arcsin \sqrt{x} \tag{4.4}$$

approximately. Note that the right side is independent of n which means

TABLE 3

THE CONTINUOUS ARC SINE DISTRIBUTION $A(x) = \frac{2}{\pi} \arcsin \sqrt{x}$

x	$A(x)$	x	$A(x)$	x	$A(x)$
0.00	0.000	0.20	0.295	0.40	0.236
0.01	0.064	0.21	0.303	0.41	0.442
0.02	0.090	0.22	0.311	0.42	0.449
0.03	0.111	0.23	0.318	0.43	0.455
0.04	0.128	0.24	0.326	0.44	0.462
0.05	0.144	0.25	0.333	0.45	0.468
0.06	0.158	0.26	0.341	0.46	0.474
0.07	0.171	0.27	0.348	0.47	0.481
0.08	0.183	0.28	0.355	0.48	0.487
0.09	0.194	0.29	0.362	0.49	0.494
				0.50	0.500
0.10	0.205	0.30	0.369		
0.11	0.215	0.31	0.376		
0.12	0.225	0.32	0.383		
0.13	0.235	0.33	0.390		
0.14	0.244	0.34	0.396		
0.15	0.253	0.35	0.403		
0.16	0.262	0.36	0.410		
0.17	0.271	0.37	0.416		
0.18	0.279	0.38	0.423		
0.19	0.287	0.39	0.429		

For $x > \frac{1}{2}$ use $A(1 - x) = 1 - A(x)$.

that table 3 suffices for all arc sine distributions of large order. (Actually the approximations are rather good even for relatively small values of n.)

We saw that, contrary to popular notions, it is quite likely that in a long coin-tossing game one of the players remains practically the whole time on the winning side, the other on the losing side. The next theorem elucidates the same phenomenon by an analysis of the fraction of the total time that the particle spends on the positive side. One feels intuitively that this fraction is most likely to be close to $\frac{1}{2}$, but the opposite is true: The possible values close to $\frac{1}{2}$ are least probable, whereas the extremes $k = 0$ and $k = n$ have the greatest probability. The analysis is facilitated by the fortunate circumstance that the theorem again involves the discrete arc sine distribution (4.1) (which will occur twice more in section 8).

Theorem 2. (*Discrete arc sine law for sojourn times.*) *The probability that in the time interval from* 0 *to* $2n$ *the particle spends* $2k$ *time units on the positive side and* $2n - 2k$ *time units on the negative side equals* $\alpha_{2k,2n}$.

(The total time spent on the positive side is necessarily even.)

Corollary.[12] *If* $0 < x < 1$, *the probability that* $\leq xn$ *time units are spent on the positive side and* $\geq (1 - x)n$ *on the negative side tends to* $\frac{2}{\pi} \arcsin \sqrt{x}$ *as* $n \to \infty$.

Examples. (*c*) From table 2 it is seen that the probability that in 20 tossings the lead never passes from one player to the other is about 0.352. The probability that the luckier player leads 16 times or more is about 0.685. (The approximation obtained from the corollary with $x = \frac{4}{5}$ is 0.590.) The probability that each player leads 10 times is only 0.06.

(*d*) Let n be large. With probability 0.20 the particle spends about 97.6 per cent of the time on the same side of the origin. In one out of ten cases the particle spends 99.4 per cent of the time on the same side.

(*e*) In example (*a*) a coin is tossed once per second for a total of 365 days. The accompanying table gives the times t_p such that with the stated

---

[12] Paul Lévy [*Sur certains processus stochastiques homogènes*, Compositia Mathematica, vol. 7 (1939), pp. 283–339] found this arc sine law for Brownian motion and referred to the connection with the coin-tossing game. A general arc sine limit law for the number of positive partial sums in a sequence of mutually independent random variables was proved by P. Erdös and M. Kac, *On the number of positive sums of independent random variables*, Bull. Amer. Math. Soc., vol. 53 (1947), pp. 1011–1020. The wide applicability of the arc sine limit law appeared at that time mysterious. The whole theory was profoundly reshaped when E. Sparre Andersen made the surprising discovery that many facets of the fluctuation theory of sums of independent random variables are of a purely combinatorial nature. [See Mathematica Scandinavica, vol. 1 (1953), pp. 263–285, and vol. 2 (1954), pp. 195–223.] The original proofs were exceedingly complicated, but they opened new avenues of research and are now greatly simplified. Theorem 2 was first proved by K. L. Chung and W. Feller by complicated methods. (See sections XII,5-6 of the first edition of this book.) Theorem 1 is new.

probability p the less fortunate player will be in the lead for a total time less than t_p.

p	t_p	p	t_p
0.9	153.95 days	0.3	19.89 days
0.8	126.10 days	0.2	8.93 days
0.7	99.65 days	0.1	2.24 days
0.6	75.23 days	0.05	13.5 hours
0.5	53.45 days	0.02	2.16 hours
0.4	34.85 days	0.01	32.4 minutes

▶

Proof of Theorem 2. Consider paths of the fixed length $2n$ and denote by $b_{2k,2n}$ the probability that exactly $2k$ sides lie above the t-axis. We have to prove that

$$b_{2k,2\nu} = \alpha_{2k,2\nu}. \tag{4.5}$$

Now (3.4) asserts that $b_{2\nu,2\nu} = u_{2\nu}$ and for reasons of symmetry we have also $b_{0,2\nu} = u_{2\nu}$. It suffices therefore to prove (4.5) for $1 \leq k \leq \nu - 1$.

Assume then that exactly $2k$ out of the $2n$ time units are spent on the positive side, and $1 \leq k \leq n - 1$. In this case a first return to the origin must occur at some epoch $2r < 2n$, and two contingencies are possible. First, the $2r$ time units up to the first return may be spent on the positive side. In this case $r \leq k \leq n - 1$, and the section of the path beyond the vertex $(2r, 0)$ has exactly $2k - 2r$ sides above the axis. Obviously the number of such paths equals $\frac{1}{2} \cdot 2^{2r} f_{2r} \cdot 2^{2n-2r} b_{2k-2r,2n-2r}$. The other possibility is that the $2r$ time units up to the first return are spent on the negative side. In this case the section beyond the vertex $(2r, 0)$ has exactly $2k$ sides above the axis, whence $n - r \geq k$. The number of such paths equals $\frac{1}{2} \cdot 2^{2r} f_{2r} \cdot 2^{2n-2r} b_{2k,2n-2r}$. Accordingly, when $1 \leq k \leq n - 1$

$$b_{2k,2n} = \frac{1}{2} \sum_{r=1}^{k} f_{2r} b_{2k-2r,2n-2r} + \frac{1}{2} \sum_{r=1}^{n-k} f_{2r} b_{2k,2n-2r}. \tag{4.6}$$

We now proceed by induction. The assertion (4.5) is trivially true for $\nu = 1$, and we assume it to be true for $\nu \leq n - 1$. Then (4.6) reduces to

$$b_{2k,2n} = \tfrac{1}{2} u_{2n-2k} \sum_{r=1}^{k} f_{2r} u_{2k-2r} + \tfrac{1}{2} u_{2k} \sum_{r=1}^{n-k} f_{2r} u_{2n-2k-2r}. \tag{4.7}$$

In view of (2.6) the first sum equals u_{2k} while the second equals u_{2n-2k}. Hence (4.5) is true also for $\nu = n$. ▶

[A paradoxical result connected with the arc sine law is contained in problem 4 of XIV,9.]

*5. CHANGES OF SIGN

The theoretical study of chance fluctuations confronts us with many paradoxes. For example, one should expect naively that in a prolonged coin-tossing game the observed number of changes of lead should increase roughly in proportion to the duration of the game. In a game that lasts twice as long, Peter should lead about twice as often. This intuitive reasoning is false. We shall show that, in a sense to be made precise, the number of changes of lead in n trials increases only as $\sqrt{n}$: in $100n$ trials one should expect only 10 times as many changes of lead as in n trials. This proves once more that the waiting times between successive equalizations are likely to be fantastically long.

We revert to random walk terminology. A *change of sign* is said to occur at epoch n if $\mathbf{S}_{n-1}$ and $\mathbf{S}_{n+1}$ are of opposite signs, that is, if the path crosses the axis. In this case $\mathbf{S}_n = 0$, and hence n is necessarily an even (positive) integer.

Theorem 1.[13] *The probability* $\xi_{r,2n+1}$ *that up to epoch* $2n + 1$ *there occur exactly* r *changes of sign equals* $2p_{2n+1,2r+1}$. *In other words*

$$\xi_{r,2n+1} = 2\mathbf{P}\{\mathbf{S}_{2n+1} = 2r + 1\}, \qquad r = 0, 1, \ldots . \tag{5.1}$$

Proof. We begin by rephrasing the theorem in a more convenient form. If the first step leads to the point (1, 1) we take this point as the origin of a new coordinate system. To a crossing of the horizontal axis in the old system there now corresponds a crossing of the line below the new axis, that is, a crossing of the level -1. An analogous procedure is applicable when $\mathbf{S}_1 = -1$, and it is thus seen that the theorem is fully equivalent to the following *proposition*: The probability that up to epoch $2n$ the level -1 is crossed exactly r times equals $2p_{2n+1,2r+1}$.

Consider first the case $r = 0$. To say that the level -1 has not been crossed amounts to saying that the level -2 has not been touched (or crossed). In this case $\mathbf{S}_{2n}$ is a non-negative even integer. For $k \geq 0$ we conclude from the basic reflection lemma of section 1 that the number of paths from $(0, 0)$ to $(2n, 2k)$ that do touch the level -2 equals the number of paths to $(2n, 2k + 4)$. The probability to reach the point

* This section is not used explicitly in the sequel.

[13] For an analogous theorem for the number of returns to the origin see problems 9-10. For an alternative proof see problem 11.

$(2n, 2k)$ without having touched the level -2 is therefore equal to $p_{2n,2k} - p_{2n,2k+4}$. The probability that the level -2 has not been touched equals the sum of the quantities for $k = 0, 1, 2, \ldots$. Most terms cancel, and we find that our probability equals $p_{2n,0} + p_{2n,2}$. This proves the assertion when $r = 0$ because

$$p_{2n+1,1} = \tfrac{1}{2}(p_{2n,0}+p_{2n,2}) \tag{5.2}$$

as is obvious from the fact that every path through $(2n + 1, 1)$ passes through either $(2n, 0)$ or $(2n, 2)$.

Next let $r = 1$. A path that crosses the level -1 at epoch $2\nu - 1$ may be decomposed into the section from $(0, 0)$ to $(2\nu, -2)$ and a path of length $2n - 2\nu$ starting at $(2\nu, -2)$. To the latter section we apply the result for $r = 0$ but interchanging the roles of plus and minus. We conclude that the number of paths of length $2n - 2\nu$ starting at $(2\nu, -2)$ and not crossing the level -1 equals the number of paths from $(2\nu, -2)$ to $(2n + 1, -3)$. But each path of this kind combines with the initial section to a path from $(0, 0)$ to $(2n + 1, -3)$. It follows that the number of paths of length $2n$ that cross the level -1 exactly once equals the number of paths from the origin to $(2n + 1, -3)$, that is, $2^{2n+1}p_{2n+1,3}$. This proves the assertion for $r = 1$.

The proposition with arbitrary r now follows by induction, the argument used in the second part of the proof requiring no change. (It was presented for the special case $r = 1$ only to avoid extra letters.) ►

An amazing consequence of the theorem is that *the probability* $\xi_{r,n}$ *of* r *changes of sign in* n *trials decreases with* r*:*

$$\xi_{0,n} \geq \xi_{1,n} > \xi_{2,n} > \cdots. \tag{5.3}$$

This means that regardless of the number of tosses, the event that the lead never changes is more probable than any preassigned number of changes.

Examples. (*a*) The probabilities x_r for exactly r changes of sign in 99 trials are as follows:

r	x_r	r	x_r
0	0.1592	7	0.0517
1	0.1529	8	0.0375
2	0.1412	9	0.0260
3	0.1252	10	0.0174
4	0.1066	11	0.0111
5	0.0873	12	0.0068
6	0.0686	13	0.0040

(*b*) The probability that in 10,000 trials no change of sign occurs is about 0.0160. The probabilities x_r for exactly r changes decrease very slowly; for $r = 10, 20, 30$ the values are $x_r = 0.0156, 0.0146$, and 0.0130. The probability that in 10,000 trials the lead changes at most 10 times is about 0.0174; in other words, one out of six such series will show not more than 10 changes of lead. ▶

A pleasing property of the identity (5.1) is that it enables us to apply the normal approximation derived in section 2. Suppose that n is large and x a fixed positive number. The probability that fewer than $x\sqrt{n}$ changes of sign occur before epoch n is practically the same as $2\mathbf{P}\{\mathbf{S}_n < 2x\sqrt{n}\}$, and according to (2.7) the last probability tends to $\mathfrak{N}(2x) - \frac{1}{2}$ as $n \to \infty$. We have thus

Theorem 2. (*Normal approximation.*) *The probability that fewer than* $x\sqrt{n}$ *changes of sign occur before epoch* n *tends to* $2\mathfrak{N}(2x) - 1$ *as* $n \to \infty$.

It follows that the *median* for the number of changes of sign is about $0.337\sqrt{n}$; this means that for n sufficiently large it is about as likely that there occur fewer than $0.337\sqrt{n}$ changes of sign than that occur more. With probability $\frac{1}{10}$ there will be fewer than $0.0628\sqrt{n}$ changes of sign, etc.[14]

6. AN EXPERIMENTAL ILLUSTRATION

Figure 4 represents the result of a computer experiment simulating 10,000 tosses of a coin; the same material is tabulated in example I, (6.*c*). The top line contains the graph of the first 550 trials; the next two lines represent the entire record of 10,000 trials the scale in the horizontal direction being changed in the ratio 1:10. The scale in the vertical direction is the same in the two graphs.

When looking at the graph most people feel surprised by the length of the intervals between successive crossings of the axis. As a matter of fact, the graph represents a rather mild case history and was chosen as the mildest among three available records. A more startling example is obtained by looking at the same graph in the *reverse* direction; that is, reversing the order in which the 10,000 trials actually occurred (see section 8). Theoretically, the series as graphed and the reversed series are equally legitimate as representative of an ideal random walk. The reversed random

[14] This approximation gives $\frac{1}{10}$ for the probability of at most 6 equalizations in 10,000 trials. This is an underestimate, the true value being about 0.112.

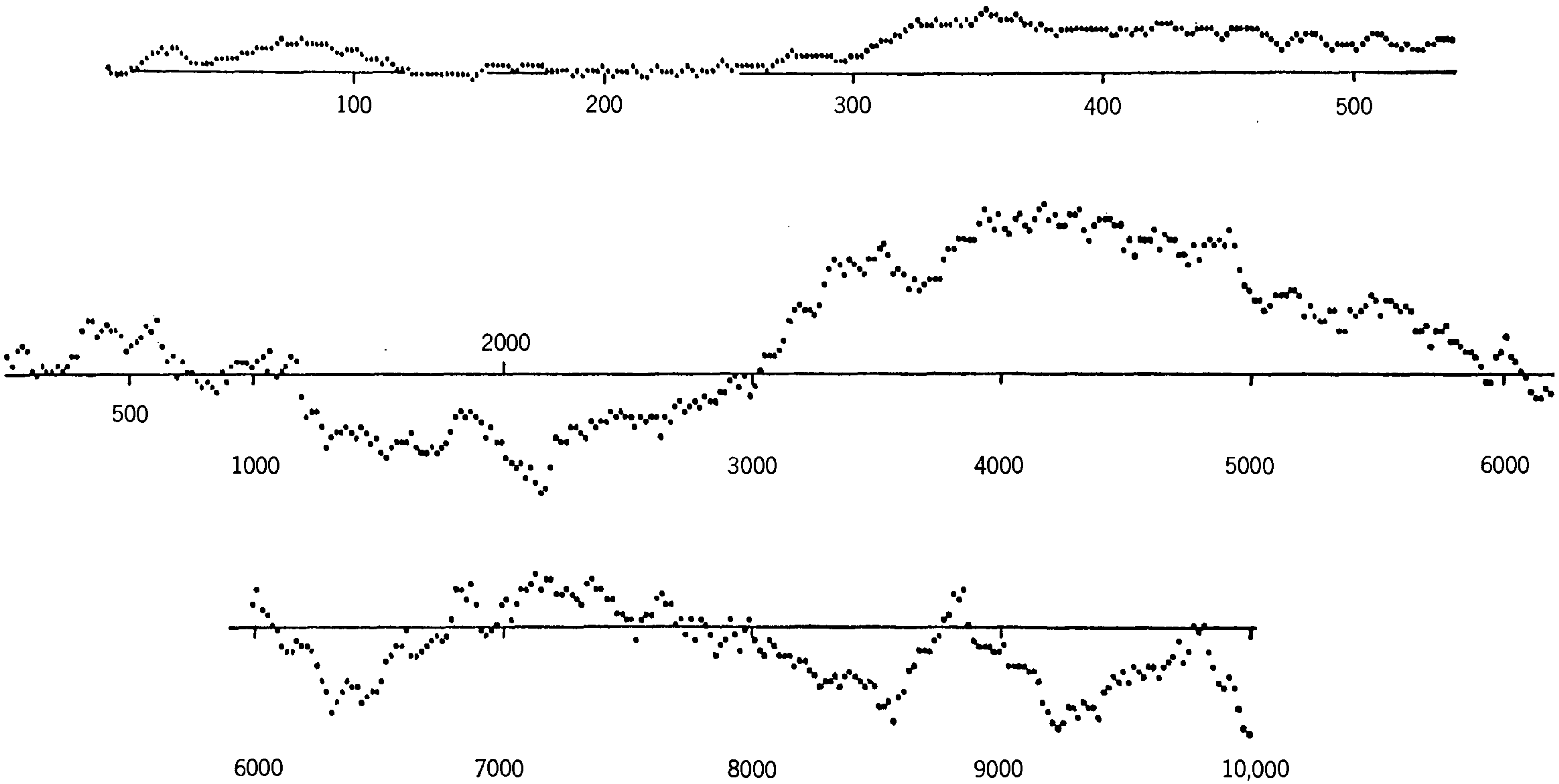

Figure 4. The record of 10,000 tosses of an ideal coin (described in section 6).

walk has the following characteristics. Starting from the origin

the path stays on the

negative side		*positive side*	
for the first	7804 *steps*	*next*	8 *steps*
next	2 *steps*	*next*	54 *steps*
next	30 *steps*	*next*	2 *steps*
next	48 *steps*	*next*	6 *steps*
next	2046 *steps*		
Total of	9930 steps	*Total of*	70 steps
Fraction of time:	0.993	*Fraction of time:*	0.007

This *looks* absurd, and yet the probability that in 10,000 tosses of a perfect coin the lead is at one side for more than 9930 trials and at the other for fewer than 70 exceeds $\frac{1}{10}$. In other words, on the average *one record out of ten will look worse than the one just described.* By contrast, the probability of a balance better than in the graph is only 0.072.

The original record of figure 4 contains 78 changes of sign and 64 other returns to the origin. The reversed series shows 8 changes of sign and 6 other returns to the origin. Sampling of expert opinion revealed that even trained statisticians expect much more than 78 changes of sign in 10,000 trials, and nobody counted on the possibility of only 8 changes of sign. Actually the probability of not more than 8 changes of sign exceeds 0.14, whereas the probability of more than 78 changes of sign is about 0.12. As far as the number of changes of sign is concerned the two records stand on a par and, theoretically, neither should cause surprise. If they seem startling, this is due to our faulty intuition and to our having been exposed to too many vague references to a mysterious "law of averages."

7. MAXIMA AND FIRST PASSAGES

Most of our conclusions so far are based on the basic lemma 3.1, which in turn is a simple corollary to the reflection principle. We now turn our attention to other interesting consequences of this principle.

Instead of paths that remain above the x-axis we consider paths that remain below the line $x = r$, that is, paths satisfying the condition

$$(7.1) \qquad S_0 < r, \qquad S_1 < r, \ldots, S_n < r.$$

We say in this case that the *maximum* of the path is $< r$. (The maximum is ≥ 0 because $S_0 = 0$.) Let $A = (n, k)$ be a point with ordinate $k \leq r$. A path from 0 to A touches or crosses the line $x = r$ if it violates the condition (7.1). By the reflection principle the number of such paths equals the number of paths from the origin to the point $A' = (n, 2r - k)$ which is the reflection of A on the line $x = r$. This proves

Lemma 1. *Let* $k \leq r$. *The probability that a path of length* n *leads to* $A = (n, k)$ *and has a maximum* $\geq r$ *equals* $p_{n,2r-k} = \mathbf{P}\{\mathbf{S}_n = 2r - k\}$.

The probability that the maximum equals r is given by the difference $p_{n,2r-k} - p_{n,2r+2-k}$. Summing over all $k \leq r$ we obtain the probability that an arbitrary path of length n has a maximum exactly equal to r. The sum is telescoping and reduces to $p_{n,r} + p_{n,r+1}$. Now $p_{n,r}$ vanishes unless n and r have the same parity, and in this case $p_{n,r+1} = 0$. We have thus

Theorem 1. *The probability that the maximum of a path of length* n *equals* $r \geq 0$ *coincides with the positive member of the pair* $p_{n,r}$ *and* $p_{n,r+1}$.

For $r = 0$ and even epochs the assertion reduces to

$$\mathbf{P}\{\mathbf{S}_1 \leq 0, \mathbf{S}_2 \leq 0, \ldots, \mathbf{S}_{2n} \leq 0\} = u_{2n}. \tag{7.2}$$

This, of course, is equivalent to the relation (3.4) which represents one version of the basic lemma. Accordingly, theorem 1 is a generalization of that lemma.

We next come to a notion that plays an important role in the general theory of stochastic processes. A *first passage through the point* $r > 0$ is said to take place at epoch n if

$$\mathbf{S}_1 < r, \ldots, \mathbf{S}_{n-1} < r, \qquad \mathbf{S}_n = r. \tag{7.3}$$

In the present context it would be preferable to speak of a first visit, but the term first passage, which originates in the physical literature, is well established; furthermore, the term visit is not applicable to continuous processes.

Obviously a path satisfying (7.3) must pass through $(n - 1, r - 1)$ and its maximum up to epoch $n - 1$ must equal $r - 1$. We saw that the probability for this event equals $p_{n-1,r-1} - p_{n-1,r+1}$, and so we have

Theorem 2. *The probability* $\varphi_{r,n}$ *that the first passage through* r *occurs at epoch* n *is given by*

$$\varphi_{r,n} = \tfrac{1}{2}[p_{n-1,r-1} - p_{n-1,r+1}]. \tag{7.4}$$

A trite calculation shows that

$$\varphi_{r,n} = \frac{r}{n}\binom{n}{\frac{n+r}{2}}2^{-n} \tag{7.5}$$

[as always, the binomial coefficient is to be interpreted as zero if $(n+r)/2$ is not an integer]. For an alternative derivation see section 8.*b*.

The distribution (7.5) is most interesting when r is large. To obtain the probability that the first passage through r occurs before epoch N we must sum $\varphi_{r,n}$ over all $n \leq N$. It follows from the normal approximation (2.7) that only those terms will contribute significantly to the sum for which r^2/n is neither very large nor very close to 0. For such terms the estimates of VII, 2 provide the approximation

$$\varphi_{r,n} \sim \sqrt{\frac{2}{\pi}} \frac{r}{\sqrt{n^3}} e^{-r^2/2n}. \tag{7.6}$$

In the summation it must be borne in mind that n must have the same parity as r. The sum is the Riemann sum to the integral in (7.7), and one is led to

Theorem 3. (*Limit theorem for first passages.*) *For fixed* t *the probability that the first passage through* r *occurs before epoch* tr^2 *tends to*[15]

$$\sqrt{\frac{2}{\pi}} \int_{1/\sqrt{t}}^{\infty} e^{-\frac{1}{2}s^2} ds = 2\left[1 - \mathfrak{N}\left(\frac{1}{\sqrt{t}}\right)\right] \tag{7.7}$$

as $r \to \infty$, *where* $\mathfrak{N}$ *is the normal distribution defined in* VII,1.

It follows that, roughly speaking, the waiting time for the first passage through r increases with the *square* of r: the probability of a first passage after epoch $\frac{9}{4}r^2$ has a probability close to $\frac{1}{2}$. It follows that there must exist points $k < r$ such that the passage from k to $k + 1$ takes a time longer than it took to go from 0 to k.

The distribution of the first-passage times leads directly to the distribution of the epoch when the particle returns to the origin for the rth time.

Theorem 4. *The probability that the rth return to the origin occurs at epoch* n *is given by the quantity* $\varphi_{r,n-r}$ *of* (7.5).

In words: An rth return at epoch n has the same probability as a first passage through r at epoch $n - r$.

Proof.[16] Consider a path from the origin to $(n, 0)$ with all sides below the axis and exactly $r - 1$ interior vertices on the axis. For simplicity we shall call such a path representative. (Figure 5 shows such a path with $n = 20$ and $r = 5$.) A representative path consists of r sections with endpoints on the axis, and we may construct 2^r different paths by assigning different signs to the vertices in the several sections (that is, by mirroring sections on the axis). In this way we obtain all paths ending with an rth return, and thus there are exactly 2^r times as many paths ending with an rth return at epoch n as there are representative paths. The theorem may

[15] (7.7) defines the so-called positive stable distribution of order $\frac{1}{2}$. For a generalization of theorem 3 see problem 14 of XIV,9.

[16] For a proof in terms of generating functions see XI,(3.17).

be therefore restated as follows: There are exactly as many representative paths of length n as there are paths of length $n - r$ ending with a first passage through r. This is so, because if in a representative path we delete the r sides whose left endpoints are on the axis we get a path of length $n - r$ ending with a first passage through r. This procedure can be reversed by inserting r sides with negative slope starting at the origin and the $r - 1$ vertices marking the first passages through $1, 2, \ldots, r - 1$. (See figure 5.) ▶

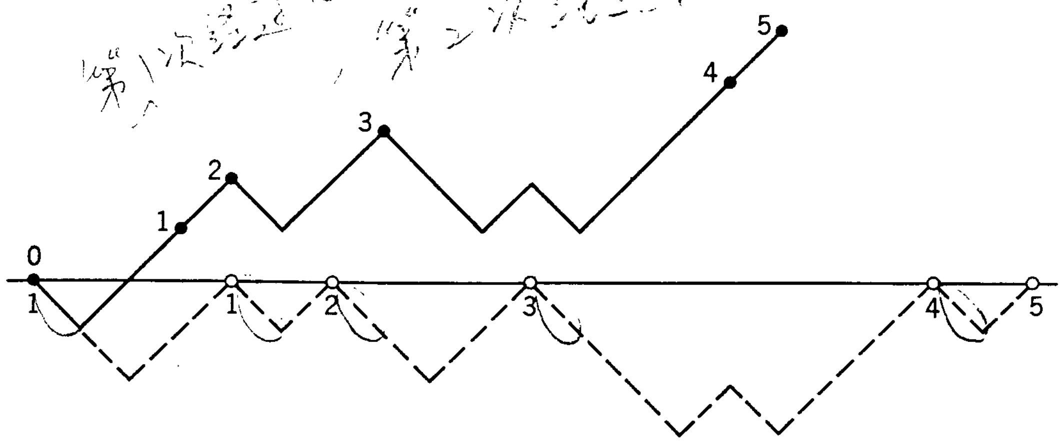

Figure 5. Illustrating first passages and returns to the origin.

It follows that the limit theorem for first returns is also applicable to rth returns as $r \to \infty$: *the probability that the rth return to the origin occurs before epoch* tr^2 *tends to the quantity* (7.7).

This result reveals another unexpected feature of the chance fluctuations in random walks. In the obvious sense the random walk starts from scratch every time when the particle returns to the origin. The epoch of the rth return is therefore the sum of r waiting times which can be interpreted as "measurements of the same physical quantity under identical conditions." It is generally believed that the average of r such observations is bound to converge to a "true value." But in the present case the sum is practically certain to be of the order of magnitude r^2, and so *the average increases roughly in proportion to* r. A closer analysis reveals that one among the r waiting times is likely to be of the same order of magnitude as the whole sum, namely r^2. In practice such a phenomenon would be attributed to an "experimental error" or be discarded as "outlier." It is difficult to see what one does not expect to see.

8. DUALITY. POSITION OF MAXIMA

Every path corresponds to a finite sequence of plus ones and minus ones, and reversing the order of the terms one obtains a new path. Geometrically the new path is obtained by rotating the given path through 180 degrees about its right endpoint, and taking the latter as origin of a new coordinate system. To every class of paths there corresponds in this way a new class of the same cardinality. If the steps of the original random walk are $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, then the steps of the new random walk are defined by

$$\mathbf{X}_1^* = \mathbf{X}_n, \ldots, \mathbf{X}_n^* = \mathbf{X}_1. \tag{8.1}$$

The vertices of the new random walk are determined by the partial sums

$$\mathbf{S}_k^* = \mathbf{X}_1^* + \cdots + \mathbf{X}_k^* = \mathbf{S}_n - \mathbf{S}_{n-k} \tag{8.2}$$

(whence $\mathbf{S}_0^* = 0$ and $\mathbf{S}_n^* = \mathbf{S}_n$). We shall refer to this as *the dual random walk.* To every event defined for the original random walk there corresponds an event of equal probability in the dual random walk, and in this way almost every probability relation has its dual. This simple method of deriving new relations is more useful than might appear at first sight. Its full power will be seen only in volume 2 in connection with general random walks and queuing theory, but even in the present context we can without effort derive some interesting new results.

To show this we shall review a few pairs of dual events, listing in each case the most noteworthy aspect. In the following list n is considered given and, to simplify language, the endpoint $(n, \mathbf{S}_n)$ of the path will be called *terminal point.* It is convenient to start from known events in the dual random walk.

(*a*) *First-passage times.* From (8.2) it is clear that the events defined, respectively, by

$$\mathbf{S}_j^* > 0, \qquad j = 1, 2, \ldots, n, \tag{8.3}$$

and

$$\mathbf{S}_n > \mathbf{S}_j, \qquad j = 0, 1, \ldots, n-1 \tag{8.4}$$

are dual to each other. The second signifies that the terminal point was not visited before epoch n. We know from (3.2) that the first event has probability $\frac{1}{2}u_{2\nu}$ when $n = 2\nu > 0$ is even; for $n = 2\nu + 1$ the probability is the same because $\mathbf{S}_{2\nu}^* > 0$ implies $\mathbf{S}_{2\nu+1}^* > 0$. Accordingly, *the probability that a first passage through a positive point takes place at epoch n equals $\frac{1}{2}u_{2\nu}$* where $\nu = \frac{1}{2}n$ or $\nu = \frac{1}{2}(n-1)$. (This is trivially true also for $n = 1$, but false for $n = 0$.) The duality principle leads us here to an interesting result which is not easy to verify directly.

(*b*) *Continuation.* In the preceding proposition the terminal point was not specified in advance. Prescribing the point r of the first passage means supplementing (8.4) by the condition $\mathbf{S}_n = r$. The dual event consists of the path from the origin to (n, r) with all intermediate vertices above the axis. The number of such paths follows directly from the reflection lemma [with $A = (1, 1)$ and $B = (n, r)$], and we get thus a new proof for (7.4).

(*c*) *Maximum at the terminal point.* A new pair of dual events is defined when the strict inequalities $>$ in (8.3) and (8.4) are changed to $\geq$. The second event occurs whenever the term $\mathbf{S}_n$ is maximal even when this maximum was already attained at some previous epoch.[17] Referring to (3.4) one sees that *the probability of this event equals* $u_{2\nu}$ where $\nu = \frac{1}{2}n$ or $\nu = \frac{1}{2}(n+1)$. It is noteworthy that the probabilities are twice the probabilities found under (*a*).

(*d*) The event that k returns to the origin have taken place is dual to the event that k visits to the terminal point occurred before epoch n. A similar statement applies to changes of sign. (For the probabilities see section 5 and problems 9–10.)

(*e*) *Arc sine law for the first visit to the terminal point.* Consider a randomly chosen path of length $n = 2\nu$. We saw under (*a*) that with probability $\frac{1}{2}u_{2\nu}$ the value $\mathbf{S}_{2\nu}$ is positive and such that no term of the sequence $\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_{2\nu-1}$ equals $\mathbf{S}_{2\nu}$. The same is true for negative $\mathbf{S}_{2\nu}$, and hence the probability that the value $\mathbf{S}_{2\nu}$ is not attained before epoch 2ν equals $u_{2\nu}$; this is also the probability of the event that $\mathbf{S}_{2\nu} = 0$ in which the terminal value is attained already at epoch 0. Consider now more generally the event that the first visit to the terminal point takes place at epoch $2k$ (in other words, we require that $\mathbf{S}_{2k} = \mathbf{S}_{2\nu}$ but $\mathbf{S}_j \neq \mathbf{S}_{2\nu}$ for $j < 2k$). This is the dual to the event that the last visit to the origin took place at epoch $2k$, and we saw in section 4 that such visits are governed by the discrete arc sine distribution. We have thus the unexpected result that *with probability* $\alpha_{2k,2\nu} = u_{2k}u_{2\nu-2k}$ *the first visit to the terminal point* $\mathbf{S}_{2\nu}$ *took place at epoch* $2\nu - 2k$ $(k = 0, 1, \ldots, \nu)$. It follows, in particular, that the epochs $2k$ and $2\nu - 2k$ are equally probable. Furthermore, very early and very late first visits are much more probable than first visits at other times.

(*f*) *Arc sine law for the position of the maxima.* As a last example of the usefulness of the duality principle we show that the results derived under (*a*) and (*c*) yield directly the probability distribution for the epochs at which the sequence $\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_n$ reaches its maximum value. Unfortunately the maximum value can be attained repeatedly, and so we must distinguish

[17] In the terminology used in chapter 12 of volume 2 we are considering a *weak ladder point* in contrast to the *strict* ladder points treated under (*a*).

between the first and the last maximum. The results are practically the same, however.

For simplicity let $n = 2\nu$ be even. The *first* maximum occurs at epoch k if

$$(8.5a) \qquad \mathbf{S}_0 < \mathbf{S}_k, \quad \ldots, \mathbf{S}_{k-1} < \mathbf{S}_k$$

$$(8.5b) \qquad \mathbf{S}_{k+1} \leq \mathbf{S}_k, \ldots \quad , \mathbf{S}_{2\nu} \leq \mathbf{S}_k.$$

Let us write k in the form $k = 2\rho$ or $k = 2\rho + 1$. According to (a) the probability of (8.5a) equals $\frac{1}{2}u_{2\rho}$, except when $k = 0$. The event (8.5b) involves only the section of the path following the epoch k and its probability obviously equals the probability that in a path of length $2\nu - k$ all vertices lie below or on the t-axis. It was shown under (c) that this probability equals $u_{2\nu-2\rho}$. Accordingly, *if* $0 < k < 2\nu$ *the probability that in the sequence* $\mathbf{S}_0, \ldots, \mathbf{S}_{2\nu}$ *the first maximum occurs at epochs* $k = 2\rho$ *or* $k = 2\rho + 1$ *is given by* $\frac{1}{2}u_{2\rho}u_{2\nu-2\rho}$. *For* $k = 0$ *and* $k = 2\nu$ *the probabilities are* $u_{2\nu}$ *and* $\frac{1}{2}u_{2\nu}$, *respectively.*

(For the *last* maximum the probabilities for the epochs 0 and 2ν are interchanged; the other probabilities remain unchanged provided k is written in the form $k = 2\rho$ or $k = 2\rho - 1$.)

We see that with a proper pairing of even and odd subscripts the position of the maxima becomes subject to the discrete arc sine distribution. Contrary to intuition the maximal accumulated gain is much more likely to occur towards the very beginning or the very end of a coin-tossing game than somewhere in the middle.

9. AN EQUIDISTRIBUTION THEOREM

We conclude this chapter by proving the theorem mentioned in connection with Galton's rank order test in example (1.b). It is instructive in that it shows how an innocuous variation in conditions can change the character of the result.

It was shown in section 4 that the number of sides lying above the x-axis is governed by the discrete arc sine distribution. We now consider the same problem but restricting our attention to paths leading from the origin to a point of the x-axis. The result is unexpected in itself and because of the striking contrast to the arc sine law.

Theorem. *The number of paths of length* $2n$ *such that* $\mathbf{S}_{2n} = 0$ *and exactly* $2k$ *of its sides lie above the axis is independent of* k *and equal to* $2^{2n}u_{2n}/(n+1) = 2^{2n+1}f_{2n+2}$. *(Here* $k = 0, 1, \ldots, n$.*)*

Proof. We consider the cases $k = 0$ and $k = n$ separately. The number of paths to $(2n, 0)$ with all sides above the x-axis equals the number of paths from $(1, 1)$ to $(2n, 0)$ which do not touch the line directly below the x-axis. By the reflection principle this number equals

$$(9.1) \qquad \binom{2n-1}{n} - \binom{2n-1}{n+1} = \frac{1}{n+1}\binom{2n}{n}.$$

This proves the assertion for $k = n$ and, by symmetry, also for $k = 0$.

For $1 \leq k \leq n - 1$ we use induction. The theorem is easily verified when $n = 1$, and we assume it correct for all paths of length less than $2n$. Denote by $2r$ the epoch of the first return. There are two possibilities. If the section of the path up to epoch $2r$ is on the positive side we must have $1 \leq r \leq k$ and the second section has exactly $2k - 2r$ sides above the axis. By the induction hypothesis a path satisfying these conditions can be chosen in

$$(9.2) \qquad 2^{2r-1}f_{2r} \cdot \frac{2^{2n-2r}}{n-r+1} u_{2n-2r} = \frac{2^{2n-2}}{r(n-r+1)} u_{2r-2}u_{2n-2r}$$

different ways. On the other hand, if the section up to the first return to the origin is on the negative side, then the terminal section of length $2n - 2r$ contains exactly $2k$ positive sides, and hence in this case $n - r \geq k$. For fixed r the number of paths satisfying these conditions is again given by (9.2). Thus the numbers of paths of the two types are obtained by summing (9.2) over $1 \leq r \leq k$ and $1 \leq r \leq n - k$, respectively. In the second sum change the summation index r to $\rho = n + 1 - r$. Then ρ runs from $k + 1$ to n, and the terms of the sum are identical with (9.2) when r is replaced by ρ. It follows that the number of paths with k positive sides is obtained by summing (9.2) over $1 \leq r \leq n$. Since k does not appear in (9.2) the sum is independent of k as asserted. Since the total number of paths is $2^{2n}u_{2n}$ this determines the number of paths in each category. (For a direct evaluation see problem 13.) ▶

An analogous theorem holds also for the position of the maxima. (See problem 14.)

10. PROBLEMS FOR SOLUTION

1. (*a*) If $a > 0$ and $b > 0$, the number of paths $(s_1, s_2, \ldots, s_n)$ such that $s_1 > -b, \ldots, s_{n-1} > -b, s_n = a$ equals $N_{n,a} - N_{n,a+2b}$.

(*b*) If $b > a > 0$ there are $N_{n,a} - N_{n,2b-a}$ paths satisfying the conditions $s_1 < b, \ldots, s_{n-1} < b, s_n = a$.

2. Let $a > c > 0$ and $b > 0$. The number of paths which touch the line $x = a$ and then lead to (n, c) without having touched the line $x = -b$ equals

$N_{n,2a-c} - N_{n,2a+2b+c}$. (Note that this includes paths touching the line $x = -b$ *before* the line $x = a$.)

3. *Repeated reflections.* Let a and b be positive, and $-b < c < a$. The number of paths to the point (n, c) which meet neither the line $x = -b$ nor $x = a$ is given by the series

$$\sum (N_{n,2k(a+b)+c} - N_{n,2k(a+b)+2a-c}),$$

the series extending over all integers k from $-\infty$ to ∞, but having only finitely many non-zero terms.

Hint: Use and extend the method of the preceding problem.

Note. This is connected with the so-called *ruin problem* which arises in gambling when the two players have initial capitals a and b so that the game terminates when the accumulated gain reaches either a or $-b$. For the connection with statistical tests, see example (1.c).

(The method of repeated reflections will be used again in problem 17 of XIV,9 and in connection with diffusion theory in volume 2; X,5.)

4. From lemma 3.1 conclude (without calculations) that

$$u_0u_{2n} + u_2u_{2n-2} + \cdots + u_{2n}u_0 = 1.$$

5. Show that

$$u_{2n} = (-1)^n\binom{-\frac{1}{2}}{n} \qquad f_{2n} = (-1)^{n-1}\binom{\frac{1}{2}}{n}.$$

Derive the identity of the preceding problem as well as (2.6) from II, (12.9).

6. Prove geometrically that there are exactly as many paths ending at $(2n + 2, 0)$ and having all interior vertices strictly above the axis as there are paths ending at $(2n, 0)$ and having all vertices above or on the axis. Therefore $\mathbf{P}\{\mathbf{S}_1 \geq 0, \ldots, \mathbf{S}_{2n-1} \geq 0, \mathbf{S}_{2n} = 0\} = 2f_{2n+2}$.

Hint: Refer to figure 1.

7. Prove lemma 3.1 geometrically by showing that the following construction establishes a one-to-one correspondence between the two classes of paths:

Given a path to $(2n, 0)$ denote its *leftmost* minimum point by $M = (k, m)$. Reflect the section from the origin to M on the vertical line $t = k$ and slide the reflected section to the endpoint $(2n, 0)$. If M is taken as origin of a new coordinate system the new path leads from the origin to $(2n, 2m)$ and has all vertices strictly above or on the axis. (This construction is due to E. Nelson.)

8. Prove formula (3.5) directly by considering the paths that never meet the line $x = -1$.

9. The probability that *before* epoch $2n$ there occur exactly r returns to the origin equals the probability that a return takes place at epoch $2n$ and is preceded by at least r returns. *Hint:* Use lemma 3.1.

10. *Continuation.* Denote by $z_{r,2n}$ the probability that exactly r returns to the origin occur up to and including epoch $2n$. Using the preceding problem show that $z_{r,2n} = p_{r,2n} + p_{r+1,2n} + \cdots$ where $p_{r,2n}$ is the probability that the rth return occurs at epoch $2n$. Using theorem 7.4 conclude that

$$z_{r,2n} = \frac{1}{2^{2n-r}} \cdot \binom{2n-r}{n}.$$

11. *Alternative derivation for the probabilities for the number of changes of sign.* Show that

$$\xi_{r,2n-1} = \frac{1}{2}\sum_{k=1}^{n-1} f_{2k}[\xi_{r-1,2n-1-2k} + \xi_{r,2n-1-2k}].$$

Assuming by induction that (5.1) holds for all epochs prior to $2n - 1$ show that this reduces to

$$\xi_{r,2n-1} = 2\sum_{1}^{n-1} f_{2k}p_{2n-2k,2r}$$

which is the probability of reaching the point $(2n, 2r)$ after a return to the origin. Considering the first step and using the ballot theorem conclude that (5.1) holds.

12. The probability that $\mathbf{S}_{2n} = 0$ and the maximum of $\mathbf{S}_1, \ldots, \mathbf{S}_{2n-1}$ equals k is the same as $\mathbf{P}\{\mathbf{S}_{2n} = 2k\} - \mathbf{P}\{\mathbf{S}_{2n} = 2k + 2\}$. Prove this by reflection.

13. In the proof of section 9 it was shown that

$$\sum_{r=1}^{n} \frac{1}{r(n-r+1)} u_{2r-2}u_{2n-2r} = \frac{1}{n+1} u_{2n}.$$

Show that this relation is equivalent to (2.6). *Hint:* Decompose the fraction.

14. Consider a path of length $2n$ with $\mathbf{S}_{2n} = 0$. We order the sides in circular order by identifying 0 and $2n$ with the result that the first and the last side become adjacent. Applying a cyclical permutation amounts to viewing the same closed path with $(k, \mathbf{S}_k)$ as origin. Show that this preserves maxima, but moves them k steps ahead. Conclude that when all $2n$ cyclical permutations are applied the number of times that a maximum occurs at r is independent of r.

Consider now a randomly chosen path with $\mathbf{S}_{2n} = 0$ and pick the place of the maximum if the latter is unique; if there are several maxima, pick one at random. This procedure leads to a number between 0 and $2n - 1$. Show that all possibilities are equally probable.