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

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

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

$$n = p + q, \qquad x = p - q. \tag{1.2}$$

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

$$N_{n,x} = \binom{p+q}{p} = \binom{p+q}{q} \tag{1.3}$$

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]

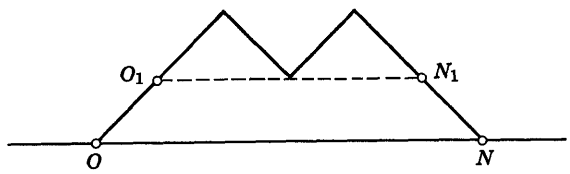

Figure 1. Illustrating positive paths. The figure shows also that there are exactly as many strictly positive paths from the origin to the point $(2n, 0)$ as there are non-negative paths from the origin to $(2n-2, 0)$.

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. ◀

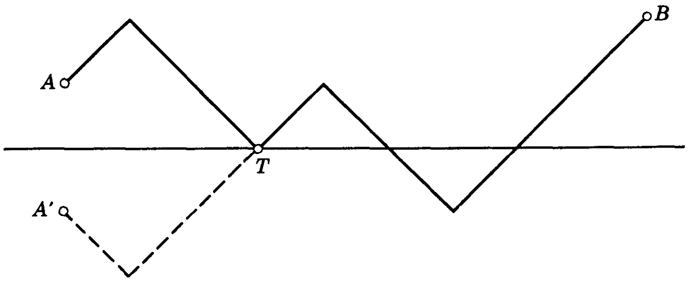

Figure 2. Illustrating the reflection principle.

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

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

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 S_j are positive, or all are negative. The two contingencies being equally probable we can restate (3.1) in the form

$$\mathbf{P}\{S_1 > 0, \ldots, S_{2n} > 0\} = \tfrac{1}{2}u_{2n}. \tag{3.2}$$

Proof. Considering all the possible values of S_{2n} it is clear that

$$\mathbf{P}\{S_1 > 0, \ldots, S_{2n} > 0\} = \sum_{r=1}^{\infty} \mathbf{P}\{S_1 > 0, \ldots, S_{2n-1} > 0, S_{2n} = 2r\} \tag{3.3}$$

(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 $\frac{1}{2}p_{2n-1,1}$. 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,

$$\mathbf{P}\{S_1 \geq 0, \ldots, S_{2n} \geq 0\} = u_{2n}. \tag{3.4}$$

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

$$\mathbf{P}\{S_1 > 0, \ldots, S_{2n} > 0\} = \tfrac{1}{2}\mathbf{P}\{S_1 \geq 0, \ldots, S_{2n-1} \geq 0\}. \tag{3.5}$$

But S_{2n-1} is an odd number, and hence $S_{2n-1} \geq 0$ implies that also $S_{2n} \geq 0$. 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

$$S_1 \neq 0, \ldots, S_{2k} \neq 0$$

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.