23. Let a and r be arbitrary positive numbers and n a positive integer. Show that

$$a(a+r)(a+2r)\cdots(a+nr) \sim Cr^{n+1}n^{n+(a/r)+\frac{1}{2}}e^{-n}. \tag{12.24}$$

[The constant C is equal to $(2\pi)^{\frac{1}{2}}/\Gamma(a/r)$.]

24. Using the results of the preceding problem, show that

$$\frac{a(a+r)(a+2r)\cdots(a+nr)}{b(b+r)(b+2r)\cdots(b+nr)} \sim \frac{\Gamma(b/r)}{\Gamma(a/r)}n^{(a-b)/r}. \tag{12.25}$$

25. Prove the following *alternative form of Stirling's formula:*

$$n! \sim (2\pi)^{\frac{1}{2}}(n+\tfrac{1}{2})^{n+\frac{1}{2}}e^{-(n+\frac{1}{2})}. \tag{12.26}$$

26. *Continuation.* Using the method of the text, show that

$$(2\pi)^{\frac{1}{2}}(n+\tfrac{1}{2})^{n+\frac{1}{2}}e^{-(n+\frac{1}{2})-1/24(n+\frac{1}{2})} < n! < (2\pi)^{\frac{1}{2}}(n+\tfrac{1}{2})^{n+\frac{1}{2}}e^{-(n+\frac{1}{2})}. \tag{12.27}$$

27. Extending Stirling's formula, prove that

$$n! \sim (2\pi)^{\frac{1}{2}}n^{n+\frac{1}{2}}\exp\left\{-n+\frac{1}{12n}-\frac{1}{360n^3}+\cdots\right\}. \tag{12.28}$$

CHAPTER III*

Fluctuations in Coin Tossing and Random Walks

This chapter serves two purposes. First, it will show that exceedingly simple methods may lead to far-reaching and important results. Second, in it we shall for the first time 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.[1]

The discussion is inserted at this place only because of its elementary character; the main topic of the book continues in chapter V. The entire book is independent of the present chapter. Some of the formulas will reappear later in connection with first passages and recurrence, but they will be derived anew by analytical methods. 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.* To facilitate such a procedure, this chapter may be read in *two versions:* the main text appears in ordinary type. Passages in small type cover additional topics (referring mainly to first passage and recurrence phenomena) and should be omitted at first reading. Section 7 contains an empirical illustration.

* 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), XIV (random walks), but the contents will not be used explicitly in the sequel.

[1] Although we are dealing formally only with coin tossing, the basic conclusions are widely applicable. In fact, E. Sparre Andersen has made the surprising discovery that many facets of the fluctuation theory of sums of independent random variables are of a purely combinatorial nature and are common to a huge class of such variables. This is true, in particular, of the two arc sine laws. See *Mathematica Scandinavica*, vol. 1 (1953), pp. 263–285, and vol. 2 (1954), pp. 195–223.

1. GENERAL ORIENTATION

A surprising wealth of information concerning chance fluctuations in general will be derived from the following inconspicuous lemma announced in 1887 by Bertrand. Similar problems of arrangements have attracted the interest of students of combinatorial analysis under the name of *ballot problems*.[2] *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* $(p - q)/(p + q)$.

In mathematical language we are here concerned with arrangements of $x = p + q$ symbols $\epsilon_1, \epsilon_2, \ldots, \epsilon_x$ consisting of p plus ones (votes for P) and q minus ones (votes for Q). The partial sum $s_k = \epsilon_1 + \epsilon_2 + \ldots + \epsilon_k$ is the number of votes by which P leads, or trails, just after the kth vote is cast. Clearly $s_x = p - q$ and

$$(1.1) \qquad s_i - s_{i-1} = \epsilon_i = \pm 1, \qquad s_0 = 0 \qquad (i = 1, 2, \ldots, x).$$

Conversely, every arrangement $\{s_1, s_2, \ldots, s_x\}$ of integers satisfying (1.1) represents a potential voting record. We shall use a geometrical terminology and represent such an arrangement by a polygonal line whose ith side has slope ϵ_i and whose ith vertex has ordinate s_i. Such lines will be called paths.

Definition. *Let $x > 0$ and y be integers. A path $\{s_1, s_2, \ldots, s_x\}$ from the origin to the point (x, y) is a polygonal line whose vertices have abscissas $0, 1, 2, \ldots, x$ and ordinates $s_0, s_1, s_2, \ldots, s_x$ satisfying* (1.1) *with $s_x = y$.*

If p among the ϵ_i are positive and q negative, then

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

An arbitrary point (x, y) can be joined to the origin by a path only if x and y are of the form (1.2). In this case the p places for the positive

[2] For the history and literature see A. Dvoretzky and T. Motzkin, A problem of arrangements, *Duke Mathematical Journal*, vol. 14 (1947), pp. 305–313. As these authors point out, most of the formally different proofs in reality use the reflection principle (lemma 1 of section 2), but without the geometric interpretation this principle loses its simplicity and appears as a curious trick. Dvoretzky and Motzkin give a new proof of great simplicity and elegance. They generalize the ballot problem by requiring that at each instant P have at least α times the votes scored by Q. This work has been continued by M. T. L. Bizley, Derivation of a new formula for the number of minimal lattice paths, etc., *The Journal of the Institute of Actuaries*, vol. 80, Part 1, No. 354 (1954), pp. 55–62.

ϵ_i can be chosen from the $x = p + q$ available places in

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

different ways. It is convenient to *define $N_{x,y} = 0$ whenever x, y are not of the form* (1.2). Then *there exist exactly $N_{x,y}$ different paths from the origin to the point* (x, y). Bertrand's ballot theorem asserts that when $y > 0$ there exist exactly $(y/x)N_{x,y}$ paths satisfying the conditions $s_1 > 0, s_2 > 0, \ldots, s_{x-1} > 0, s_x = y$. It will be proved in section 2.

Example. Figure 1 exhibits a path to the point $N_1 = (5,1)$. There exist ten such paths of which two satisfy the conditions $s_i > 0$. The path in the graph is $\{1, 2, 1, 2, 1\}$, and the other is $\{1, 2, 3, 2, 1\}$.

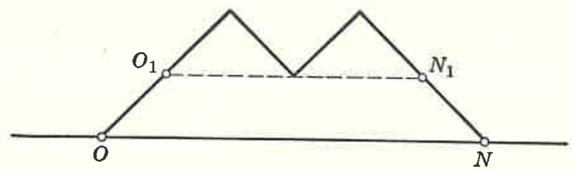

FIGURE 1. Illustrating positive paths and the proof of theorem 2 in section 2.

We can draw the most interesting conclusions from the ballot theorem if we drop the convention that the terminal point (x, y) of the path be fixed in advance. There exist 2^n different paths from the origin to points (n, y) with an arbitrary ordinate y. As explained in section 3, these 2^n paths may be taken to represent the 2^n possible outcomes of the ideal experiment consisting in n successive tossings of a perfect coin. 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*, that is, the excess of the accumulated number of heads over tails.

If $s_n = 0$, the net gain at the conclusion of the nth trial is zero: there exists a *tie*. Ties occur so infrequently that they do not affect the picture, but repeated references to them are disturbing. We shall therefore agree to say that *at the nth trial Peter leads if either $s_n > 0$ or $s_n = 0$ but $s_{n-1} > 0$* (i.e., in case of a tie that player leads who led at the preceding trial). "Peter leads at the nth trial" is but a description for "the nth side of the path is above the x-axis."

The ballot theorem refers to paths situated entirely above the x-axis, that is, to games in which the lead never changes. This topic may be pursued further by investigating how often the lead is likely to change for an arbitrary path. In this connection we reach conclusions that play havoc with our intuition. It is generally expected that in a prolonged series of coin tossings Peter should lead about half the time and Paul the other half. This is entirely wrong, however. *In 20,000 tossings it is about 88 times more probable that Peter leads in all 20,000 trials than that each player leads in 10,000 trials.* In general, the lead changes at such infrequent intervals that intuition is defied. No matter how long the series of tossings, the *most probable number of changes of lead is zero;* exactly one change of lead is more probable than two, two changes are more probable than three, etc. In short, if a modern educator or psychologist were to describe the long-run case histories of individual coin-tossing games, he would classify the majority of coins as maladjusted. If many coins are tossed n times each, a surprisingly large proportion of them will leave one player in the lead almost all the time; and in very few cases will the lead change sides and fluctuate in the manner that is generally expected of a well-behaved coin.

This is a sample of the conclusions to be drawn from the first arc sine law (see section 5 and the illustration in section 7). E. Sparre Andersen has shown that this law has a wide field of applicability, and the situation here described for coin tossings is typical for chance fluctuations involving cumulative effects. Most stochastic processes in physics, economics, and education are of this nature, and our findings should serve as a warning to those who are prone to discern secular trends and deviations from average norms.

The same situation may be viewed from a somewhat different angle. If the coin tossing proceeds at a uniform rate, common sense expects that, with due allowance for chance fluctuations, a two-day game should produce twice as many ties as a one-day game. In other words, we expect intuitively the number of ties to increase roughly in proportion to the duration of the game. Paradoxically this is not so: *The number of ties increases about as the square root of time.* In 10,000 tossings the median number of ties is 67, but in 1,000,000 tossings it increases only to 674; the typical "*wavelength*" increases from about 150 to about 1500. *The average wavelength increases with time* (sections 6 and 8). The formulas on which these conclusions are based play an important role for first passage and recurrence times in general random walks and diffusion theory.

Theorem 3 of section 2 stands apart from the remainder and is not used elsewhere. It concerns a variant of the ballot problem for the case where the two candidates score the same number, n, of votes. Then P leads an even number, $2k$, of times and Q leads in the remaining $2n - 2k$ trials. Again we have the false intuition that each candidate is likely to lead about half the time, that is, we expect $2k$ to be close to n. Actually, if the ballot ended in a tie $n:n$, *the $n + 1$ possible divisions of leads (namely $2n:0$, $2n-2:2$, $2n-4:4$, $\ldots$, $2:2n-2$, $0:2n$) have the same probability*

$(n + 1)^{-1}$. This result stands in a marked contrast to the situation described above where the end result was not prescribed in advance; there the extreme divisions $2n:0$ and $0:2n$ are most probable.

It has been pointed out by J. L. Hodges [3] that this theorem has statistical applications to *rank-order* tests. We illustrate this point by the

Example. Suppose that a quantity (e.g. the height of plants) is measured on each of n treated subjects and also on each of n control subjects, obtaining measurements $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$. To fix ideas, suppose that each group is arranged in decreasing order: $a_1 > a_2 > \ldots$ and $b_1 > b_2 > \ldots$. Let us combine the two sequences and write the $2n$ letters $a_1, \ldots, b_n$ in decreasing order. The resulting arrangement of n letters a and n letters b may be interpreted as the record of a ballot in which each candidate received n votes. For an extremely successful treatment all the a's should precede the b's; a completely ineffectual treatment should produce a random order. In our arrangement the a's lead exactly $2k$ times if k different a's precede the b's of same rank, that is, if the inequality $a_i > b_i$ holds for exactly k subscripts. Assuming randomness, the probability that this happens equals $1/(n + 1)$ and therefore the probability that the a's lead $2k$ times *or more* is $(n - k + 1)/(n + 1)$. The classical example for this argument (used qualitatively without knowledge of the theoretical probabilities) is due to Galton who used it in 1876 for data referred to him by Charles Darwin. In his example $2n$ was 30 and the a's were in the lead 26 times. Galton concluded that the treatment *was* efficient, but on the hypothesis of mere randomness even an ineffectual treatment would produce 26 or more leads in three out of sixteen similar experiments. This shows that a qualitative analysis may be a valuable supplement to our rather shaky intuition. (For related tests based on the theory of runs see chapter II, section 5*a*.)

2. PROBLEMS OF ARRANGEMENTS

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 x-axis is meant the point $A' = (a, -\alpha)$.* (See figure 2.) A *path from A to B* is defined as in section 1, with A playing the role of the origin.

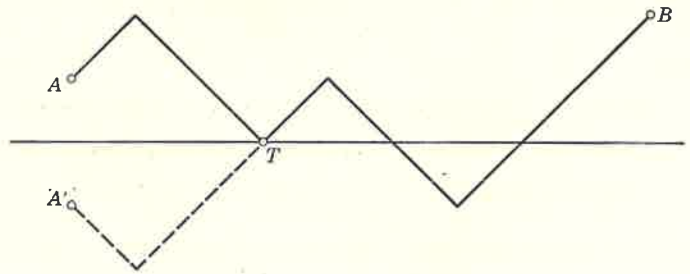

FIGURE 2. Illustrating the reflection principle.

[3] Galton's rank-order test, *Biometrika*, vol. 42 (1955), pp. 261–262.

Lemma.[4] (*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 x-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 path leading from A' to B and having $T = (t, 0)$ as its first vertex on the x-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 as have a vertex on the x-axis. The lemma is proved.

Theorem 1. (*Ballot theorem.*) *Let $x > 0$, $y > 0$; the number of paths $\{s_1, s_2, \ldots, s_x = y\}$ from the origin to (x, y) such that $s_1 > 0$, $s_2 > 0, \ldots, s_x > 0$ equals $(y/x)N_{x,y}$.*

Proof. Since $s_1 = \pm 1$, we have $s_1 = 1$ for each admissible path. It follows that there exist as many admissible paths as there are paths leading from the point $(1, 1)$ to (x, y) which neither touch nor cross the x-axis. By the last lemma the number of such paths equals

$$(2.1)\qquad N_{x-1,y-1} - N_{x-1,y+1} = \binom{p+q-1}{p-1} - \binom{p+q-1}{q-1}$$

$$= \frac{p-q}{p+q}\binom{p+q}{p} = \frac{y}{x}N_{x,y}.$$

The Duality Principle. Almost every theorem on paths can be reformulated to obtain a formally different theorem. Consider $\{s_1, \ldots, s_x\}$ and the path obtained from it by reversing the order of the ϵ_i, that is, the path $\{s_1^*, s_2^*, \ldots, s_x^*\}$ where $s_1^* = \epsilon_x$, $s_2^* = \epsilon_x + \epsilon_{x-1}$, $s_3^* = \epsilon_x + \epsilon_{x-1} + \epsilon_{x-2}, \ldots$ or

$$(2.2)\qquad s_0^* = 0,\quad s_1^* = s_x - s_{x-1},\quad \ldots,\quad s_i^* = s_x - s_{x-i},\quad \ldots,\quad s_x^* = s_x.$$

The two paths (1.1) and (2.2) are congruent and are obtained from each other by a rotation through 180 degrees; they join the same endpoints. *To each theorem on paths there corresponds a dual theorem obtained by applying it to the reversed path* (2.2).

For example, the ballot theorem gives us the number of reversed paths $\{s_1^*, \ldots, s_x^*\}$ joining the origin to (x, y) such that $s_i^* > 0$ for $i = 1, 2, \ldots, x$. But this is

[4] The probability literature attributes this method to D. André (1887). The text reduces it to a lemma on random walks. The classical difference equations of random walks (chapter XIV) closely resemble differential equations, and the reflection principle (even a stronger form of it) is familiar in that theory under the name of *Lord Kelvin's method of images.*

the same as $s_x > s_{x-i}$ for $i = 1, 2, \ldots, x-1$ and hence we have as an alternative form of the ballot theorem

Theorem 1*. *The number of paths $\{s_1, s_2, \ldots, s_x\}$ from $(0, 0)$ to (x, y) such that $s_1 < s_x$, $s_2 < s_x, \ldots, s_{x-1} < s_x$ (where $s_x = y > 0$) equals $(y/x)N_{x,y}$.*

Geometrically speaking, theorem 1 is concerned with paths whose left endpoint is the lowest vertex, whereas the dual theorem 1* refers to paths whose last vertex is highest. (See figure 3.) Theorem 1* has implications for first-passage times in random walks.

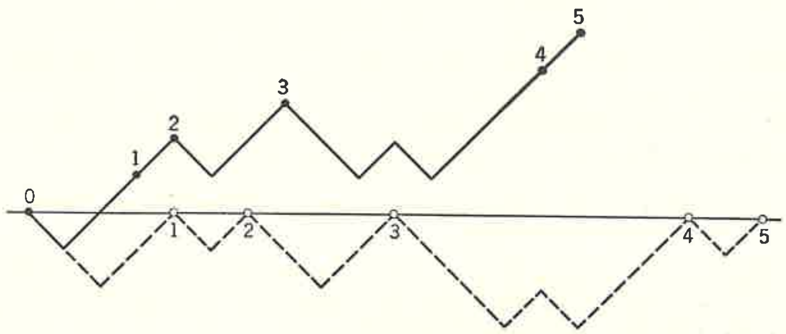

FIGURE 3. Illustrating first passages and returns to the origin.

We turn to a study of paths joining the origin to a point $N = (2n, 0)$ of the x-axis (an odd vertex on the x-axis is impossible). Put for abbreviation

$$(2.3)\qquad L_{2n} = \frac{1}{n+1}\binom{2n}{n}.$$

Theorem 2. *Among the $\binom{2n}{n}$ paths joining the origin to the point $2n$ of the x-axis there are*

(*a*) *exactly L_{2n-2} paths such that*

$$(2.4)\qquad s_1 > 0,\quad s_2 > 0,\quad \ldots,\quad s_{2n-1} > 0,\qquad (s_{2n} = 0)$$

(*b*) *exactly L_{2n} paths such that*

$$(2.5)\qquad s_1 \geq 0,\quad s_2 \geq 0,\quad \ldots,\quad s_{2n-1} \geq 0,\qquad (s_{2n} = 0).$$

(*That is, there are as many paths to $2n$ with all inner vertices above the x-axis as there are paths to $2n - 2$ with no vertex below the x-axis.*)

Proof. (See figure 1.) Each path satisfying condition (2.4) passes through the point $N_1 = (2n-1, 1)$ and by theorem 1 the number of paths to N_1 such that $s_1 > 0, \ldots, s_{2n-2} > 0$ equals

$$\frac{1}{2n-1}\binom{2n-1}{n-1} = \frac{1}{n}\binom{2n-2}{n-1} = L_{2n-2}. \tag{2.6}$$

This proves (*a*). Again, let a path satisfy condition (2.4). Omitting the first and the last side we get a path that joins the point $O_1 = (1, 1)$ to $N_1 = (2n-1, 1)$ and at the same time is such that all its vertices lie on or above the line $y = 1$. Translating the origin to O_1, we get a path from the new origin to the point N_1 (which has the new coordinates $2n-2$ and 0), none of whose vertices lies below the new x-axis. We have thus established a one-to-one correspondence between such paths and all paths satisfying (2.4), and the theorem is proved.

As explained in section 1 the following theorem stands apart from the remainder and will not be used in the sequel.

Theorem 3.[5] *Let $L_{2k,2n}$ be the number of paths from the origin to the point $2n$ of the x-axis such that $2k$ of its sides lie above the x-axis and $2n - 2k$ below ($k = 0, 1, \ldots, n$). Then $L_{2k,2n} = L_{2n}$, independently of k.*

Proof. The assertion is trivially true for $n = 1$ and we assume by induction that $L_{2k,2\nu} = L_{2\nu}$ for $\nu = 1, 2, \ldots, n-1$ and $0 \leq k \leq \nu$. We propose to count the number of paths $\{s_1, s_2, \ldots, s_{2n} = 0\}$ with exactly $2k$ sides above the x-axis. First assume $1 \leq k \leq n-1$. Such a path crosses the x-axis and we denote by $2r$ the abscissa of its *first* vertex on the x-axis. We have then to consider two classes of paths.

A path of the first class is positive between 0 and $2r$, and its section between $2r$ and $2n$ contains exactly $2k - 2r$ sides above the axis. Here $k \geq r$. By theorem 2(*a*) there exist L_{2r-2} paths $\{s_1, \ldots, s_{2r-1}, s_{2r} = 0\}$ with $s_1 > 0, \ldots, s_{2r-1} > 0$, and by the induction hypothesis there exist $L_{2k-2r,2n-2r} = L_{2n-2r}$ paths joining $(2r, 0)$ to $(2n, 0)$ and having $2k - 2r$ sides above the x-axis. Accordingly, there exists a total of $L_{2r-2}L_{2n-2r}$ paths of this class.

A path of the second class is negative between 0 and $2r$; its section between $2r$ and $2n$ then contains $2k$ sides above the x-axis. By the argument above there exist again $L_{2r-2}L_{2n-2r}$ paths of this class, but this time $n - r \geq k$.

It follows that for $k = 1, \ldots, n-1$

$$L_{2k,2n} = \sum_{r=1}^{k} L_{2r-2}L_{2n-2r} + \sum_{r=1}^{n-k} L_{2r-2}L_{2n-2r}. \tag{2.7}$$

By changing the summation index to $\rho = n - r + 1$, the terms of the second

[5] First proved by complicated analytical methods by K. L. Chung and W. Feller, Fluctuations in coin tossing, *Proceedings National Academy of Sciences USA*, vol. 35 (1949), pp. 605–608 (see also the first edition of the present book, chapter XII, problem 4). An elegant combinatorial proof was given by J. L. Hodges (see footnote 3).

sum become $L_{2r-2}L_{2n-2r} = L_{2\rho-2}L_{2n-2\rho}$ with ρ running from $k + 1$ to n. Thus

$$L_{2k,2n} = \sum_{\rho=1}^{n} L_{2\rho-2}L_{2n-2\rho}, \tag{2.8}$$

which is *independent of k*.

A path with all $2n$ sides above the x-axis is a path of the sort described in theorem 2(*b*), and hence $L_{2n,2n} = L_{2n}$. For reasons of symmetry we have also $L_{0,2n} = L_{2n}$. The total number of paths from the origin to $(2n, 0)$ being $(n + 1)L_{2n}$, it follows that $L_{2k,2n} = L_{2n}$ for $k = 0, 1, \ldots, n$.

As a *corollary* we find the identity

$$L_{2n} = \sum_{\rho=1}^{n} L_{2\rho-2}L_{2n-2\rho}. \tag{2.9}$$

For a direct analytic verification see section 8(*a*).

3. RANDOM WALKS AND COIN TOSSING

In a sequence of N tossings of an ideal coin let $\epsilon_k = +1$ if the kth trial results in heads and $\epsilon_k = -1$ otherwise. Then $s_k = \epsilon_1 + \epsilon_2 + \ldots + \epsilon_k$ is the cumulative excess of heads over tails at the conclusion of the kth trial. In classical betting language s_k is "Peter's accumulated net gain." Each possible outcome of the N successive tossings is represented by a path of N sides starting at the origin, and conversely each such path may be taken as representing the outcome of N tossings.

This consideration leads us to *take for our sample space the aggregate of the 2^N paths $\{s_1, \ldots, s_N\}$ starting at the origin and to attribute probability 2^{-N} to each.*

An event such as "heads at the first two trials" must be interpreted as the aggregate of all sequences starting with $s_1 = 1$, $s_2 = 2$. There are 2^{N-2} such sequences and the probability of this event is therefore 2^{-2}, as is proper. More generally, if $k < N$ there exist exactly 2^{N-k} different paths $\{s_1, s_2, \ldots, s_N\}$ such that their first k vertices lie on a preassigned path $\{s_1, s_2, \ldots, s_k\}$. It follows that *an event determined by the outcome of the first $k < N$ trials has a probability independent of N*. In practice, therefore, the number N plays no role, provided it is sufficiently large. Conceptually and formally it is best to consider each finite sequence of tossings as the beginning of a potentially infinite sequence, but this would lead us into non-denumerable sample spaces. We shall therefore consider finite sequences with N larger than the number of trials occurring in the formulas; except for this we shall be permitted, and be glad, to forget about N.

For the probabilistic background and the connection with related topics it is desirable to supplement the geometric language by an alternative terminology. We imagine the coin tossings performed at a uniform rate, so that the nth trial occurs at *time n*. Peter may mark his cumulative gain at all times by an indicator which we shall call "particle." This particle, then, moves on a vertical axis starting from the origin. It moves at times 1, 2, ... one unit step upward if the coin lands heads, one unit step downward if the coin lands tails. We say that *the particle performs a symmetric random walk.* (The physicist takes it as the simplest model for one-dimensional diffusion; see chapter XIV.)

At time n the position of the particle is the point s_n of the vertical axis. *The path* $\{s_1, s_2, \ldots, s_N\}$ *represents the space-time diagram of the random walk,* the x-axis playing the role of the time axis.

Guided by this background we introduce the following

Terminology. *We shall say that at time n there takes place:*

A return to the origin if $s_n = 0$.

A first return to the origin if

$$s_1 \neq 0, \quad s_2 \neq 0, \quad \ldots, \quad s_{n-1} \neq 0, \quad s_n = 0. \tag{3.1}$$

A first passage through $r > 0$ *if*

$$s_1 < r, \quad s_2 < r, \quad \ldots, \quad s_{n-1} < r, \quad s_n = r. \tag{3.2}$$

A second, third, ... return to the origin and a first passage through $r < 0$ are defined in an obvious way. Note that passages through the origin can take place only at *even* times, and we shall frequently restrict the formulas to even times. In betting language a return to the origin represents an *equalization of the accumulated numbers of heads and tails.* (Figure 3 exhibits two paths in which the first passages and returns to the origin, respectively, are marked; the second path has the peculiarity of keeping to the negative side.)

4. REFORMULATION OF THE COMBINATORIAL THEOREMS

In the following sections we shall use the notations

$$u_{2n} = \binom{2n}{n} 2^{-2n}, \qquad n = 0, 1, 2, \ldots \tag{4.1}$$

and

$$f_0 = 0, \qquad f_{2n} = \frac{1}{2n} u_{2n-2}, \qquad n = 1, 2, \ldots \tag{4.2}$$

It is easily verified that

$$f_{2n} = u_{2n-2} - u_{2n}, \qquad n = 1, 2, \ldots. \tag{4.3}$$

Theorem 1. *For each* $n \geq 1$:

$$u_{2n} = \mathbf{P}\{s_{2n} = 0\} \tag{4.4}$$

$$u_{2n} = \mathbf{P}\{s_1 \neq 0, s_2 \neq 0, \ldots, s_{2n} \neq 0\} \tag{4.5}$$

$$u_{2n} = \mathbf{P}\{s_1 \geq 0, s_2 \geq 0, \ldots, s_{2n} \geq 0\} \tag{4.6}$$

or in words: The three events, (a) a return to the origin takes place at time $2n$, *(b) no return occurs up to and including time* $2n$, *and (c) the path is non-negative between* 0 *and* $2n$, *have the common probability* u_{2n}.

Furthermore,

$$f_{2n} = \mathbf{P}\{s_1 \neq 0, s_2 \neq 0, \ldots, s_{2n-1} \neq 0, s_{2n} = 0\} \tag{4.7}$$

$$f_{2n} = \mathbf{P}\{s_1 \geq 0, s_2 \geq 0, \ldots, s_{2n-2} \geq 0, s_{2n-1} < 0\} \tag{4.8}$$

that is: the two events (a) the first return to the origin takes place at time $2n$, *and (b) the first passage through* -1 *occurs at time* $2n - 1$, *have the common probability* f_{2n}.

Proof. As was observed in section 3 it suffices to consider the sample space of paths of the fixed length $2n$. By (1.3) there exist $\binom{2n}{n}$ paths joining the origin to the point $(2n, 0)$, and this proves (4.4).

By theorem 2(*a*) in section 2 there exist L_{2n-2} paths joining the origin to $(2n, 0)$ such that $s_1 > 0, \ldots, s_{2n-1} > 0$. Therefore there are twice as many paths satisfying the condition in (4.7), and the corresponding probability is $2L_{2n-2} \cdot 2^{-2n} = f_{2n}$. Theorem 2(*b*) in section 2 implies in the same way (4.8).

The probability that no zero occurs up to and including time $2n$ equals one minus the probability of a first return to the origin at a time $\leq 2n$. Using (4.7) this difference is

$$\begin{aligned} 1 - f_2 - f_4 - \ldots - f_{2n} &= \\ 1 - (1 - u_2) - (u_2 - u_4) - \ldots &- (u_{2n-2} - u_{2n}) = u_{2n} \end{aligned} \tag{4.9}$$

which proves (4.5). Similarly, the right side in (4.6) equals one minus the probability of a first passage through -1 before time $2n$, and using (4.8) this difference is again given by (4.9). This accomplishes the proof.

Corollary. *It follows that for* $n \geq 1$

$$u_{2n} = \sum_{r=1}^{n} f_{2r} u_{2n-2r}. \tag{4.10}$$

Proof. If a return to the origin takes place at time $2n$, then the *first* return must take place at some time $2r \le 2n$. We have just seen that the number of paths from the origin to $(2n, 0)$ with the first return to the origin taking place at time $2r \le 2n$ equals $2^{2r}f_{2r}\cdot 2^{2n-2r}u_{2n-2r}$. Summing over r, we get equation (4.10). (For a direct analytic proof see section 8(*a*). In chapter XIII, section 3, we shall see that (4.10) is a special case of the basic equation for recurrent events.)

Theorem 1* in section 2 enumerates the paths in which a first passage through y occurs at time x. The sum $x + y$ must be even, and for our purposes it is convenient to put $x = 2n - y$. The content of theorem 1* may then be restated as follows.

Theorem 2. *The probability that a first passage through $y > 0$ takes place at time $2n - y$ is given by*

$$f_{2n}^{(y)} = \frac{y}{2n-y}\binom{2n-y}{n}2^{-2n+y}, \qquad n \ge y > 0. \tag{4.11}$$

The simplicity with which the duality principle delivered this important formula as a direct consequence of the ballot theorem is truly remarkable. A direct analytic derivation of (4.11) is difficult and requires special tricks.

In principle, the probabilities $f_{2n}^{(y)}$ can be calculated by induction on y. A path of length $2n - y - 1$ in which a first passage through $y + 1$ occurs at the terminal point may be decomposed into two segments (see figure 3 for $y = 4$). The first segment is the path from the origin up to the point of the first passage through y; it occurs at some time $2\nu - y < 2n - y - 1$. This section is followed by the second, a section of length $2n - 2\nu - 1$ in which the terminal endpoint is the only one lying above the left endpoint. In other words, if its left endpoint is taken as the origin, the second section represents a path with a first passage through 1 at the endpoint. By definition there exist $2^{2\nu-y}f_{2\nu}^{(y)}$ sections of the first type and $2^{2n-2\nu-1}f_{2n-2\nu}^{(1)}$ of the second, and any two can be combined to give a path with first passage through $y + 1$ at time $2n - y - 1$. Therefore

$$f_{2n}^{(y+1)} = \sum_{\nu=y}^{n-1} f_{2\nu}^{(y)}f_{2n-2\nu}^{(1)}, \qquad n \ge y + 1. \tag{4.12}$$

Formula (4.8) states that a first passage through -1 (and hence also through $+1$) at time $2n - 1$ has probability f_{2n}, that is,

$$f_{2n}^{(1)} = f_{2n} \qquad n \ge 1. \tag{4.13}$$

Equations (4.12) and (4.13) determine recursively all $f_{2n}^{(y)}$, but it is not easy to verify that (4.11) satisfies (4.12), and it is not at all clear how the explicit formula (4.11) could be derived from (4.12).

Formulas (4.12)–(4.13) permit a novel conclusion. We see from (4.13) that $f_{2n}^{(1)}$ is the probability that the first return to zero occurs at time $2n$. Forgetting about the preceding theorem, let us now define $f_{2n}^{(y)}$ as the probability that the yth return to zero takes place at time $2n$. The argument used in the last proof applies without change: Splitting a path from the origin to the $(y+1)$st return into the initial

section leading to the yth return and the terminal section between the yth and the $(y+1)$st return, we see again that (4.12) holds. Since this relation uniquely determines all $f_{2n}^{(y)}$ we have

Theorem 3. *The probability that the yth return to zero takes place at time $2n$ is given by* (4.11).

Alternative geometric proof. Consider a path leading from the origin to a first passage through y at time $2n - y$. (Figure 3 exhibits the case $y = 5$, $2n - y = 15$.) Construct a new path by inserting into this path y new sides each of slope -1 and having left endpoints, respectively, at the origin and the $y - 1$ vertices at which a first passage through 1, 2, ..., $y-1$ takes place. The new path, say $\{\sigma_1, \sigma_2, \ldots, \sigma_{2n}\}$, has length $2n$. Clearly $\sigma_1 \le 0, \ldots, \sigma_{2n-1} \le 0$, $\sigma_{2n} = 0$, and exactly $y - 1$ interior vertices lie on the x-axis. Conversely, each path $\{\sigma_1, \ldots, \sigma_{2n}\}$ with this property is obtained, in the manner described, from a path with first passage through y at time $2n - y$. If $f_{2n}^{(y)}$ is defined as in theorem 2, we see that there exist exactly $2^{2n-y}f_{2n}^{(y)}$ paths $\{\sigma_1, \ldots, \sigma_{2n}\}$ such that $\sigma_i \le 0$, $\sigma_{2n} = 0$, and exactly $y - 1$ interior vertices lie on the x-axis. Such a path consists of y sections with endpoints on the x-axis, and we can produce 2^y different paths by changing the signs of all σ_i of one or more such sections. In this way we obtain *all* paths of length $2n$ with $s_{2n} = 0$ and exactly $y - 1$ inner vertices on the x-axis, and their number is therefore $2^{2n}f_{2n}^{(y)}$, as asserted.

5. PROBABILITY OF LONG LEADS: THE FIRST ARC SINE LAW

We shall say that *the particle spends the time from $k - 1$ to k on the positive side if the kth side of its path lies above the x-axis*, that is, if at least one of the two vertices s_{k-1} and s_k is positive (in which case the other is positive or zero). In the betting terminology this means that at both the $(k-1)$st and the kth trial Peter's accumulated gain was non-negative.

The paradoxical properties of the paths mentioned in section 1 will be derived from the following

Theorem 1.[6] *Let $p_{2k,2n}$ be 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. Then*

$$p_{2k,2n} = u_{2k}u_{2n-2k}. \tag{5.1}$$

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

Proof. The probability that the particle keeps to the positive side during the entire time interval from 0 to $2n$ is given by formula (4.6),

[6] First proved by complicated analytical methods by K. L. Chung and W. Feller (see footnote 5 and the first edition of the present book, chapter XII, sections 5 and 6). The theorem was suggested by the work of E. Sparre Andersen (see footnote 1).

and we see that $p_{2n,2n} = u_{2n}$ as asserted. For reasons of symmetry we have also $p_{0,2n} = u_{2n}$, and it remains only to prove (5.1) for $1 \leq k \leq n-1$. For that purpose we repeat the argument which led to (2.7). A particle that keeps for $2k > 0$ time units to the positive side and for $2n - 2k > 0$ time units to the negative side necessarily passes through zero. Let $2r$ be the moment of its *first* return to zero. Then the path belongs to one of the following two classes.

In the first class, up to time $2r$ the particle stays on the positive side, and during the time interval from $2r$ to $2n$ it spends exactly $2k - 2r \geq 0$ time units on the positive side. There exist $2^{2r}f_{2r}$ paths of length $2r$ which return to the origin for the first time at $2r$, and half of them keep to the positive side. Furthermore, by definition, there are $2^{2n-2r}p_{2k-2r,2n-2r}$ paths of length $2n - 2r$ starting at $(2r, 0)$ and having exactly $2k - 2r$ sides above the x-axis. Thus the total number of paths of length $2n$ in the first class equals

$$\tfrac{1}{2}\cdot 2^{2r}f_{2r}\cdot 2^{2n-2r}p_{2k-2r,2n-2r} = 2^{2n-1}f_{2r}p_{2k-2r,2n-2r}.$$

In the second class, from 0 to $2r$ the particle keeps to the negative side, and between $2r$ and $2n$ it spends $2k$ time units on the positive side. Here $2k \leq 2n - 2r$ and the argument above shows that the number of paths in this class equals $2^{2n-1}f_{2r}p_{2k,2n-2r}$.

It follows that for $1 \leq k \leq n-1$

$$(5.2)\qquad p_{2k,2n} = \tfrac{1}{2}\sum_{r=1}^{k} f_{2r}p_{2k-2r,2n-2r} + \tfrac{1}{2}\sum_{r=1}^{n-k} f_{2r}p_{2k,2n-2r}.$$

Suppose now by induction that $p_{2k,2\nu} = u_{2k}u_{2\nu-2k}$ for $\nu = 1, 2, \ldots, n-1$ (this relation being trivially true for $\nu = 1$). Then formula (5.2) reduces to

$$(5.3)\qquad p_{2k,2n} = \tfrac{1}{2}\,u_{2n-2k}\sum_{r=1}^{k} f_{2r}u_{2k-2r} + \tfrac{1}{2}\,u_{2k}\sum_{r=1}^{n-k} f_{2r}u_{2n-2k-2r}.$$

In view of equation (4.10), the first sum equals u_{2k} and the second equals u_{2n-2k} and therefore (5.1) holds.

We feel intuitively that the *fraction k/n of the total time spent on the positive side* is most likely to be close to $\frac{1}{2}$. However, the opposite is true: *The possible values close to $\frac{1}{2}$ are least probable and the extreme values $k/n = 0$ and $k/n = 1$ have the greatest probability.* This assertion can be verified using a ratio test on (5.1).

Table 1 illustrates the paradox. In betting terminology it reveals the startling fact that in $2n = 20$ tossings of a perfect coin *with probability 0.3524 the less fortunate player will never be in the lead.* In most cases (with probability 0.5379) the accumulated gain of the less fortunate player will be positive just once or never. By contrast, an equal division 10:10 of the leads has a probability of only 0.0606.

TABLE 1

DISTRIBUTION OF LEADS IN 20 TOSSES OF A COIN

	$k = 0$ $k = 20$	$k = 2$ $k = 18$	$k = 4$ $k = 16$	$k = 6$ $k = 14$	$k = 8$ $k = 12$	$k = 10$
$p_{k,20} =$ $P_{k,20} =$	0.1762 0.3524	0.0927 0.5379	0.0736 0.6851	0.0655 0.8160	0.0617 0.9394	0.0606 1

$p_{k,20} = u_k u_{20}$ is the probability that k sides of the path are above the axis, i.e., "Peter leads during exactly k out of the 20 trials."
$P_{k,20}$ is the probability that one of the players is in the lead for at least k trials, the other for at most $20 - k$ trials.

Formula (5.1), although exact, is not very revealing, and it is preferable to replace it by a simpler approximation. An easy application of Stirling's formula II(9.1) shows that $u_{2n}(\pi n)^{\frac{1}{2}} \to 1$ as $n \to \infty$. [This is the content of problem II(12.20).] It follows that

$$(5.4)\qquad p_{2k,2n} \sim \frac{1}{\pi k^{\frac{1}{2}}(n-k)^{\frac{1}{2}}}$$

where the ratio of the two sides tends rapidly to unity as $k \to \infty$ and $n - k \to \infty$. The probability that the fraction k/n of the time spent on the positive side lies between $\frac{1}{2}$ and α ($\frac{1}{2} < \alpha < 1$) is given by

$$(5.5)\qquad \sum_{\frac{1}{2}n<k<\alpha n} p_{2k,2n} \sim \frac{1}{\pi n}\sum_{\frac{1}{2}n<k<\alpha n}\left\{\frac{k}{n}\left(1-\frac{k}{n}\right)\right\}^{-\frac{1}{2}}$$

On the right side we recognize the Riemann sum approximating the integral

$$(5.6)\qquad \pi^{-1}\int_{\frac{1}{2}}^{\alpha}\frac{dx}{\{x(1-x)\}^{\frac{1}{2}}} = 2\pi^{-1}\arcsin\alpha^{\frac{1}{2}} - \tfrac{1}{2}.$$

For reasons of symmetry the probability that $k/n \le \frac{1}{2}$ tends to $\frac{1}{2}$ as $n \to \infty$. Adding this probability to (5.5), we get

Theorem 2.[7] (*The first arc sine law.*) *For fixed* α $(0 < \alpha < 1)$ *and* $n \to \infty$ *the probability that the fraction* k/n *of time spent on the positive side be* $<\alpha$ *tends to*

$$\pi^{-1}\int_0^{\alpha}\frac{dx}{\{x(1-x)\}^{\frac{1}{2}}} = 2\pi^{-1}\arcsin \alpha^{\frac{1}{2}}. \tag{5.7}$$

In practice formula (5.7) provides an excellent approximation even for values of n as small as 20. The integrand in (5.7) is represented by a U-shaped curve tending to infinity at the endpoints 0 and 1. This

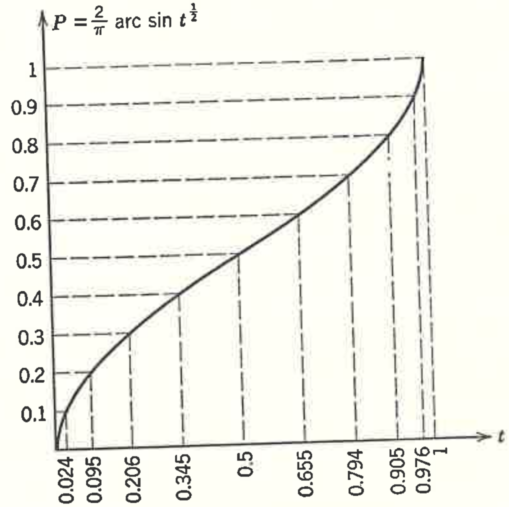

FIGURE 4. The arc sine law.

shows in a striking fashion that the fraction of time spent on the positive side is much more likely to be close to zero or to one than to the "expected" or "normal" value $\frac{1}{2}$. Figure 4 will reveal:

[7] Paul Lévy (Sur certains processus stochastiques homogènes, *Compositio Mathematica*, vol. 7 (1939), pp. 283–339) found the arc sine law for certain continuous diffusion processes and referred to the connection with the coin-tossing game. A general arc sine 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, *Bulletin of the American Mathematical Society*, vol. 53 (1947), pp. 1011–1020. It was E. Sparre Andersen who discovered the combinatorial nature of the arc sine law and its validity for general classes of random variables.

With probability 0.20 *the particle stays for about* 97.6 *per cent of the time on the same side of the origin. In one out of* 10 *cases the particle will spend* 99.4 *per cent of the time on the same side.* Another illustration is given in table 2.

TABLE 2

ILLUSTRATING THE ARC SINE LAW

p	t_p
0.9	153.95 days
.8	126.10 days
.7	99.65 days
.6	75.23 days
.5	53.45 days
.4	34.85 days
.3	19.89 days
.2	8.93 days
.1	2.24 days
.05	13.5 hours
.02	2.16 hours
.01	32.4 minutes

A coin is tossed once per second for a total of 365 days; let **Z** be the fraction of time during which the less fortunate player is in the lead. Then t_p is a number such that the event $\mathbf{Z} < t_p$ has probability p, approximately.

This table shows the probability p that the less fortunate player will be in the lead for a total of less than t_p days of a full year. Using, for example, the significance level $p = 0.05$ dear to statisticians, we see that *in one out of* 20 *cases the more fortunate player will be in the lead for more than* 364 *days and* 10 *hours.* Few people will believe that a perfect coin will produce preposterous sequences in which no change of lead occurs for millions of trials in succession, and yet this is what a good coin will do rather regularly.

In the next section we shall treat another aspect of the same phenomenon, and in section 7 we shall illustrate the theory by empirical material.

6. THE NUMBER OF RETURNS TO THE ORIGIN

The explanation of the arc sine law lies in the fact that frequently enormously many trials are required before the particle returns to the origin. Geometrically speaking, the path crosses the x-axis very rarely.

We feel intuitively that if Peter and Paul toss a coin for a long time $2n$, *the number of ties* (moments when the cumulative scores are equal) *should be roughly proportional to* $2n$. *But this is not so.* Actually the number of ties increases in probability only as $(2n)^{\frac{1}{2}}$; that is, with increasing duration of the game the frequency of ties decreases rapidly, and the "waves" increase in length. In analyzing this situation we shall consider the number of returns to zero. It should be borne in mind that the number of times when the particle actually crosses from the positive side into the negative or conversely is roughly *one-half* the number of returns.

Theorem 1. *Let $z_{2n}^{(r)}$ be the probability that up to and including time $2n$ the particle returns to zero exactly r times. Then*

$$z_{2n}^{(r)} = \frac{1}{2^{2n-r}}\binom{2n-r}{n}, \qquad n \geq 1. \tag{6.1}$$

In particular $z_{2n}^{(0)} = z_{2n}^{(1)} = u_{2n}$ and

$$z_{2n}^{(0)} = z_{2n}^{(1)} > z_{2n}^{(2)} > z_{2n}^{(3)} > \ldots. \tag{6.2}$$

In words (6.2) states that, *independently of the duration $2n$ of the game, it is more likely that no return or exactly one return to zero has occurred than any other number.*

Proof. We recall that by formulas (4.4) and (4.5) there exist exactly as many paths of length 2ν with no return to zero as there are paths with a return to zero at the last step. Consider now paths of length $2n$ in which the rth *and last* return occurred at some time $2n - 2\nu < 2n$. The section of length 2ν starting at this last return can be chosen in as many ways as we can choose an alternative section starting at the same point $(2n-2\nu, 0)$ of the x-axis and leading to $(2n, 0)$. In other words: *The probability that exactly r returns to zero occur before time $2n$ equals the probability that a return occurs at time $2n$ and that it is preceded by at least r returns.* By theorem 3 of section 4 this means that

$$z_{2n}^{(r)} = f_{2n}^{(r)} + f_{2n}^{(r+1)} + f_{2n}^{(r+2)} + \cdots. \tag{6.3}$$

with $f_{2n}^{(y)}$ given by (4.11). It is easily verified that

$$f_{2n}^{(y)} = \frac{1}{2^{2n-y}}\binom{2n-y}{n} - \frac{1}{2^{2n-y-1}}\binom{2n-y-1}{n}, \tag{6.4}$$

and adding for $y = r, r+1, \ldots$ we get equation (6.1) as asserted. The assertion (6.2) being a trivial consequence, the theorem is proved.

It is again desirable to replace the exact formula (6.1) by a simpler approximation. For that purpose we rewrite (6.1) in the form

$$z_{2n}^{(r)} = u_{2n}\frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{r-1}{n}\right)}{\left(1-\frac{1}{2n}\right)\left(1-\frac{2}{2n}\right)\cdots\left(1-\frac{r-1}{2n}\right)}. \tag{6.5}$$

As was pointed out in the proof of the arc sine law, we have $u_{2n}(\pi n)^{\frac{1}{2}} \to 1$ as $n \to \infty$. From the Taylor expansion of the logarithm, II(8.10), we see that $\log(1 - \nu/n)$ may be approximated by $-\nu/n$ with an error of the order of magnitude $(\nu/n)^2$. It follows that with an error of the magnitude r^3/n^2 we have the approximation

$$\log\{z_{2n}^{(r)}\pi^{\frac{1}{2}}n^{\frac{1}{2}}\} \approx -\frac{1}{2n}\sum_{\nu=1}^{r-1}\nu \approx -\frac{r^2}{4n} \tag{6.6}$$

or

$$z_{2n}^{(r)} \approx \pi^{-\frac{1}{2}}n^{-\frac{1}{2}}e^{-r^2/4n}. \tag{6.7}$$

The probability of fewer than k returns, namely $z_{2n}^{(0)} + z_{2n}^{(1)} + \ldots + z_{2n}^{(k-1)}$, is thus approximated by a Riemann sum to the integral over $\pi^{-1}e^{-\frac{1}{4}x^2}$ extended from 0 to k/n, the relative (or percentage) error involved being of the order of magnitude k^3/n^2. We have thus

Theorem 2. *For each fixed $\alpha > 0$ the probability that up to and including time $2n$ the particle returns to the origin fewer than $\alpha(2n)^{\frac{1}{2}}$ times tends as $n \to \infty$ to* [8]

$$f(\alpha) = (2/\pi)^{\frac{1}{2}}\int_0^{\alpha} e^{-\frac{1}{2}s^2}\, ds. \tag{6.8}$$

In particular, the probability that there occur fewer than $0.6745(2n)^{\frac{1}{2}}$ returns is, for large n, approximately $\frac{1}{2}$.

In chapter VII, section 1, the reader will find a table of the normal distribution function $\Phi(\alpha) = \frac{1}{2}\{1 + f(\alpha)\}$; from it the values $f(\alpha)$ may be obtained using $f(\alpha) = 2\{\Phi(\alpha) - \frac{1}{2}\}$ for $\alpha > 0$.

Let a coin be tossed 10,000 times: with probability $\frac{1}{2}$ there will be fewer than 68 returns to zero, of which only one-half represent actual changes of the lead. In other words, with probability $\frac{1}{2}$ the mean duration of a "wave" between two consecutive changes of lead is about 300. For 1,000,000 tossings the median number of returns has increased only by a factor 10, and the *mean* duration of a wave has increased to about 3000. The longer the series of trials, the rarer the returns to zero and the longer the waves.

The probability that in 10,000 *tossings of a coin the lead never changes* is about 0.0085, and with the same probability there will be fewer than 10 changes of lead in 1,000,000 tossings.

7. AN EXPERIMENTAL ILLUSTRATION

Figure 5 represents the result of an experiment simulating 10,000 tosses of a coin; it is the material tabulated in example I(6.*c*). The top line contains the graph of the first 550 trials, and the next two lines represent the entire record of 10,000 trials on a smaller scale in the x-direction. The scale in the y-direction is the same on the two graphs.

When looking at the graph most people feel surprised by the length of the waves between successive crossings of the x-axis (i.e., successive changes of lead). Nevertheless, the graph represents a comparatively mild case history and was chosen as the mildest among three available records. The reader is asked to look at the *same graph in the reverse direction*, that is, to take the terminal point as origin. [Analytically, the reversed path is given by (2.2).] Theoretically, the series as graphed and the reversed series are equivalent, and each represents a

[8] Readers acquainted with the central limit theorem are warned that the number of returns is *not* normally distributed. In (6.8) there appears a *truncated* normal distribution with mean $(2/\pi)^{\frac{1}{2}}$ and variance $1 - 2/\pi$.

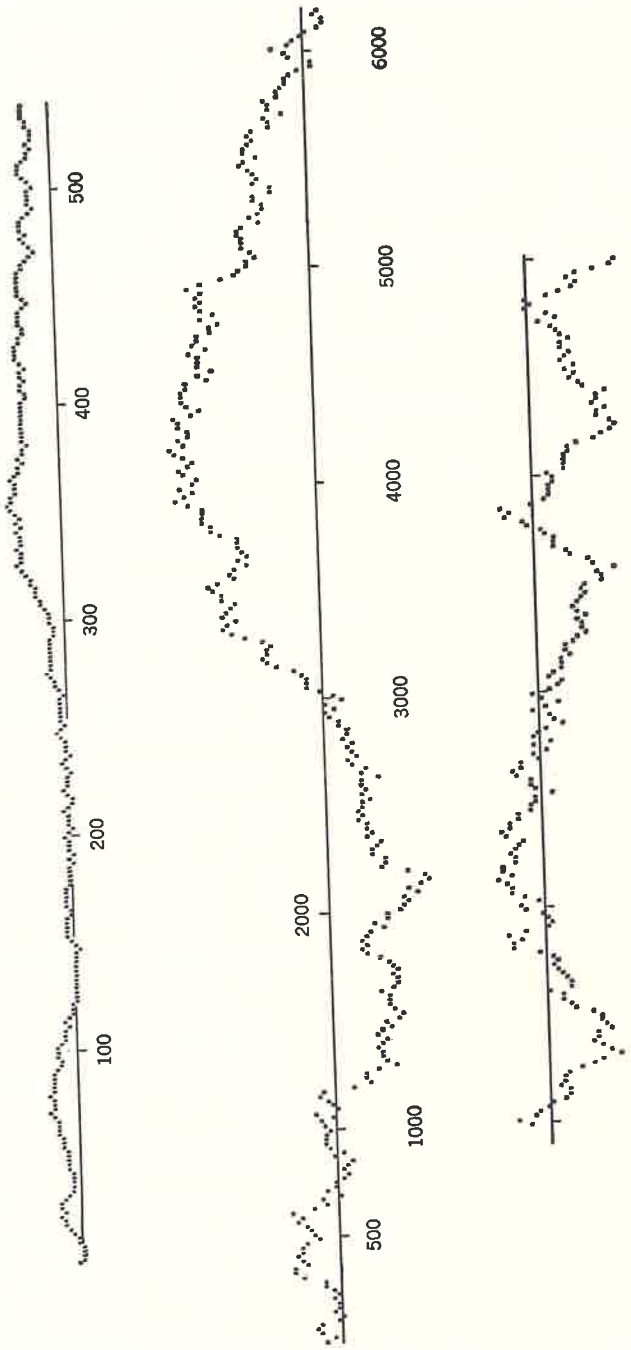

FIGURE 5. The record of 10,000 tosses of an ideal coin (described in section 7).

random walk. The reversed random walk has the following characteristics. Starting from the origin

the "particle" stays at the

	negative side		*positive side*
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: 0.9930		*fraction of time:* 0.007	

This *looks* absurd, and yet the probability that in 10,000 tosses of a perfect coin the lead is on one side for more than 9930 trials and at the other for fewer than 70 trials is slightly greater than 0.1. In other words, on the average *more than one record out of ten will look worse than the one just described.* By contrast, the probability of a record showing a better balance of leads than that of figure 5 is smaller, namely about 0.072.

The record of figure 5 contains 142 returns to the origin among which there are 78 actual changes of lead. The reversed series described above contains 14 returns of which 8 are changes of lead. Sampling of expert opinion has revealed that even trained statisticians feel that 142 equalizations in 10,000 tosses of a coin is a surprisingly small number, and 14 appears quite out of bounds. Actually *the probability of more than* 140 *equalizations is about* 0.157 *while the probability of fewer than* 14 *equalizations is about* 0.115. Thus, contrary to intuition, finding only 14 equalizations is not surprising at all; as far as the number of changes of lead is concerned, the reversed series stands on a par with the original series of figure 5.

8. MISCELLANEOUS COMPLEMENTS

(a) Analytical Verification of Identities

It is easily verified that

$$u_{2n} = (-1)^n \binom{-\frac{1}{2}}{n}, \qquad f_{2n} = (-1)^{n+1}\binom{\frac{1}{2}}{n}. \tag{8.1}$$

The basic identity (4.10) can now be regarded as a special case of equation II(12.9) for $a = \frac{1}{2}$, $b = -\frac{1}{2}$. The same formula shows in addition that $\sum_{r=0}^{n} u_{2r}u_{2n-2r} = 1$.

Formula (2.8) may be rewritten in terms of f_{2k} instead of L_{2k+2} and reduces to the special case of II(12.9) for $a = b = \frac{1}{2}$. Alternatively, formula (2.8) may be derived from (4.10) using the identity $nr^{-1}(n-r)^{-1} = r^{-1} + (n-r)^{-1}$.

(b) The Position of the Maxima: The Second Arc Sine Law

We shall say that *the path* $\{s_1, s_2, \ldots, s_x\}$ *has its first maximum at the place* k *if*

$$s_k > 0, \quad s_k > s_1, \quad \ldots, \quad s_k > s_{k-1}, \quad s_k \geq s_{k+1}, \quad \ldots, \quad s_k \geq s_x. \tag{8.2}$$

In particular, the first maximum is at the place 0 if $s_j \leq 0$ for $1 \leq j \leq x$. By formula (4.6) the probability that a path of length $x = 2n$ has its first maximum at 0 equals u_{2n}. It follows that also for a path of length $x = 2n - 1$ the probability of the first maximum at 0 equals u_{2n}.

The event "first maximum at the *last* place" is the same as $s_j < s_x$ for $j = 0, 1, \ldots, x-1$. For the reversed path (2.2) this means $s_1^* > 0$, $s_2^* > 0, \ldots, s_x^* > 0$, and the probability of this is given by (4.5), namely $\frac{1}{2}u_{2n}$ for $x = 2n$ and also for $x = 2n + 1$.

A path of length $2n$ with a first maximum at k consists of two sections: The initial section has its first maximum at the last, or kth, place, and the second section has its first maximum at the initial, or zero-th, place. Conversely, any two sections with the stated properties may be combined to give a path with its first maximum at the kth place. We have thus the

Theorem. *The probability that a path of length* $2n$ *has its first maximum at the place* ν *equals*

$$\tfrac{1}{2}u_{2k}u_{2n-2k} \quad \begin{array}{lll} \text{if} & \nu = 2k & (k = 1, 2, \ldots, n) \\ \text{or} & \nu = 2k+1 & (k = 0, 1, \ldots, n-1) \end{array} \tag{8.3}$$

and u_{2n} *if* $\nu = 0$.

The remarkable fact is that *the probability of finding the first maximum at either* $2k$ *or* $2k + 1$ *equals the probability* $p_{2k,2n}$ in (5.1) that the particle spends $2k$ out of $2n$ time units on the positive side. It follows that the arc sine approximation applies and we can conclude that there is a *strong tendency for the maxima to occur near one or the other of the endpoints.*

The surprising circumstance that the probability distribution $\{p_{2k,2n}\}$ of leads and the distribution of the position of the maxima are practically the same is no peculiarity of the coin-tossing game. An analogous theorem has been proved by E. Sparre Andersen for a large class of random variables, and the combinatorial basis of his proof is similar to the argument used above.

(c) A Limit Theorem for First Passages and Returns to the Origin[9]

The estimates used in section 6 may be used to show that for fixed $y > 0$ the probability $f_{2n}^{(y)}$ of (4.11) satisfies the asymptotic relation

$$f_{2n}^{(y)} \sim \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{y}{(2n-y)^{\frac{3}{2}}} e^{-\frac{1}{2}y^2/(2n-y)}, \tag{8.4}$$

the sign $\sim$ indicating that the ratio of the two sides tends to unity as $n \to \infty$. The methods employed for the limit theorems in sections 5 and 6 now lead to the following conclusion: *The probability that the* y*th return to zero (or the first passage through* y*) takes place before time* ty^2 *tends, with increasing* y*, to* $1 - f(t^{-\frac{1}{2}})$ *with* $f(\alpha)$ *defined in equation* (6.8).

It follows that *with probability near* $\frac{1}{2}$ *the* y*th return to zero will occur after time* $(2.21\ldots)y^2$, so that *the average time between consecutive returns is bound to increase roughly linearly with* y. This should come as a surprise to physicists accustomed to taking the average of y "measurements on the same quantity" as approximation to the "true" value. In the present case a closer analysis reveals that in all likelihood one among the y measurements will be of the same order of magnitude as the whole sum, namely y^2.

---

[9] This is theorem 3 of chapter XII, section 5, in the first edition. Advanced readers are advised that $1 - f(t^{-\frac{1}{2}})$ is the so-called positive stable distribution of order $\frac{1}{2}$.

arise, and therefore (2.4) is to be replaced by the simpler equation

$$P_0(t+h) = P_0(t)(1-\lambda h) + o(h), \tag{2.5}$$

which leads to

$$P'_0(t) = -\lambda P_0(t). \tag{2.6}$$

From (2.6) and $P_0(0) = 1$ we get $P_0(t) = e^{-\lambda t}$. Substituting this $P_0(t)$ into (2.4) with $n = 1$, we get an ordinary differential equation for $P_1(t)$. Since $P_1(0) = 0$, we find easily that $P_1(t) = \lambda t e^{-\lambda t}$, in agreement with the Poisson distribution (1.1). Proceeding in the same way, we find successively all terms of (1.1).

3. THE PURE BIRTH PROCESS

In the Poisson process the probability of a change during $(t, t+h)$ is independent of the number of changes during $(0, t)$. The simplest generalization consists of dropping this assumption. Assume instead that, when n changes occur during $(0, t)$, the probability of a new change during $(t, t+h)$ equals $\lambda_n h$ plus terms of smaller order of magnitude than h; the single constant λ characterizing the process is replaced by the sequence $\lambda_0, \lambda_1, \lambda_2, \ldots$.

It is convenient to introduce a more flexible terminology. Instead of saying that n changes occur during $(0, t)$, we shall say that *the system is in state* E_n. A new change then becomes a *transition* $E_n \to E_{n+1}$. In a pure birth process transitions from E_n are possible only to E_{n+1}. Such a process is characterized by the following

Postulates. *If at time t the system is in state* E_n $(n = 0, 1, 2, \ldots)$, *then the probability that during* $(t, t+h)$ *a transition to* E_{n+1} *occurs equals* $\lambda_n h + o(h)$; *the probability of any other change is* $o(h)$.

The salient feature of this assumption is that the time which the system spends in any particular state plays no role; there are sudden changes of state but no aging as long as the system remains within a single state.

Again let $P_n(t)$ be the probability that at time t the system is in state E_n. The functions $P_n(t)$ satisfy a system of differential equations which can be derived by the argument of the preceding section, with the only change that (2.2) is replaced by

$$P_n(t+h) = P_n(t)(1-\lambda_n h) + P_{n-1}(t)\lambda_{n-1}h + o(h). \tag{3.1}$$

In this way we get the *basic system of differential equations*

$$\begin{aligned} P'_n(t) &= -\lambda_n P_n(t) + \lambda_{n-1}P_{n-1}(t) \qquad (n \geq 1), \\ P'_0(t) &= -\lambda_0 P_0(t). \end{aligned} \tag{3.2}$$

We can calculate $P_0(t)$ first and then, by recursion, all $P_n(t)$. If the state of the system represents the number of changes during $(0, t)$, then the initial state is E_0 so that $P_0(0) = 1$ and hence $P_0(t) = e^{-\lambda_0 t}$. However, the system need not start from state E_0 [see example (3.*b*)]. If at time zero the system is in E_i, then we have

$$P_i(0) = 1, \qquad P_n(0) = 0 \qquad \text{for } n \neq i. \tag{3.3}$$

These *initial conditions* uniquely determine the solution $\{P_n(t)\}$ of (3.2). (In particular, $P_0(t) = P_1(t) = \ldots = P_{i-1}(t) = 0$.) Explicit formulas for $P_n(t)$ have been derived independently by many authors but are of no interest to us. It is easily verified that for arbitrarily prescribed λ_n the system $\{P_n(t)\}$ has all required properties, except that under certain conditions $\Sigma P_n(t) < 1$. This phenomenon will be discussed in section 4.

Examples. (*a*) *Radioactive transmutations.* A radioactive atom, say uranium, may by emission of particles or γ-rays change to an atom of a different kind. Each kind represents a possible state of the system, and as the process continues, we get a succession of transitions $E_0 \to E_1 \to E_2 \to \ldots \to E_m$. According to accepted physical theories, the probability of a transition $E_n \to E_{n+1}$ remains unchanged as long as the atom is in state E_n, and this hypothesis is expressed by our starting supposition. The differential equations (3.2) therefore describe the process (a fact well known to physicists). If E_m is the terminal state from which no further transitions are possible, then $\lambda_m = 0$ and the system (3.2) terminates with $n = m$. (For $n > m$ we get automatically $P_n(t) = 0$.)

(*b*) *The Yule process.* Consider a population of members which can (by splitting or otherwise) give birth to new members but cannot die. Assume that during any short time interval of length h each member has probability $\lambda h + o(h)$ to create a new one; the constant λ determines the rate of increase of the population. If there is no interaction among the members and at time t the population size is n, then the probability of an increase during $(t, t+h)$ is $n\lambda h + o(h)$. The probability $P_n(t)$ that the population numbers exactly n elements therefore satisfies (3.2) with $\lambda_n = n\lambda$, that is,

$$P'_n(t) = -n\lambda P_n(t) + (n-1)\lambda P_{n-1}(t) \qquad (n \geq 1). \tag{3.4}$$

If i is the population size at time $t = 0$, then the initial conditions (3.3) apply. It is easily verified that for $n \geq i$ the solution is given by

$$P_n(t) = \binom{n-1}{n-i} e^{-i\lambda t}(1-e^{-\lambda t})^{n-i} \tag{3.5}$$

and, of course, $P_n(t) = 0$ for $n < i$. This distribution is a special case of the negative binomial distribution: using the definition VI(8.1) we may rewrite (3.5) as $P_n(t) = f(n-i; i, e^{-\lambda t})$. It follows [cf. example IX(3.c)] that the population size at time t is the sum of i independent random variables each having the distribution obtained from (3.5) on replacing i by 1. These i variables represent the progenies of the i original members of our population.

This type of process was first studied by Yule[7] in connection with the mathematical theory of evolution. The population consists of the species within a genus, and the creation of a new element is due to mutations. The assumption that each species has the same probability of throwing out a new species neglects the difference in species sizes. Since we have also neglected the possibility that a species may die out, formula (3.5) can be expected to give only a crude approximation. Furry[8] used the same model to describe a process connected with cosmic rays, but again the approximation is rather crude. The differential equations (3.4) apply strictly to a population of particles which can split into exact replicas of themselves, provided, of course, that there is no interaction among particles.

* 4. DIVERGENT BIRTH PROCESSES

The solution $\{P_n(t)\}$ of the infinite system of differential equations (3.2) subject to initial conditions (3.3) can be calculated inductively, starting from $P_i(t) = e^{-\lambda_i t}$. The distribution $\{P_n(t)\}$ is therefore uniquely determined. From the familiar formulas for solving linear differential equations it follows also that $P_n(t) \geq 0$. The only question

[7] G. Udny Yule, A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S., *Philosophical Transactions of the Royal Society, London*, Series B, vol. 213 (1924), pp. 21–87. Yule does not introduce the differential equations (3.4) but derives $P_n(t)$ by a limiting process similar to the one used in chapter VI, section 5, for the Poisson process. Much more general, and more flexible, models of the same type were devised and applied to epidemics and population growth in an unpretentious and highly interesting paper by Lieutenant Colonel A. G. M'Kendrick, Applications of mathematics to medical problems, *Proceedings Edinburgh Mathematical Society*, vol. 44 (1925), pp. 1–34. It is very unfortunate that this remarkable paper passed practically unnoticed. In particular, it was unknown to the present author when he introduced various stochastic models for population growth in Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung, *Acta Biotheoretica*, vol. 5 (1939), pp. 11–40.

[8] On fluctuation phenomena in the passage of high-energy electrons through lead, *Physical Reviews*, vol. 52 (1937), p. 569.

* This section treats a special topic and may be omitted.

left open is whether $\{P_n(t)\}$ is an honest probability distribution, that is, whether or not

$$\Sigma P_n(t) = 1 \tag{4.1}$$

for all t. We shall see that this is not always so: if the coefficients λ_n increase sufficiently fast, then it may happen that

$$\Sigma P_n(t) < 1. \tag{4.2}$$

At first sight this possibility appears surprising and, perhaps, disturbing, but it finds a ready explanation. The left side in (4.2) may be interpreted as the probability that during time t only a *finite number* of changes takes place. Accordingly, the difference between the two sides in (4.2) accounts for the possibility of infinitely many changes, or a sort of explosion. For a better understanding of this phenomenon let us compare our probabilistic model of growth with the familiar deterministic approach.

The quantity λ_n in (3.2) could be called the average rate of growth at a time when the population size is n. For example, in the special case (3.4) we have $\lambda_n = n\lambda$, so that the average rate of growth is proportional to the actual population size. If growth is not subject to chance fluctuations and has a rate of increase proportional to the instantaneous population size, then $x(t)$ varies in accordance with the deterministic differential equation

$$\frac{dx(t)}{dt} = \lambda x(t). \tag{4.3}$$

It follows that at time t the population size is

$$x(t) = ie^{\lambda t}, \tag{4.4}$$

where $i = x(0)$ is the initial population size. The connection between (3.4) and (4.3) is not purely formal. It is readily seen that (4.4) actually gives the expected value of the distribution (3.5), so that (4.3) describes the expected population size, whereas (3.4) takes account of chance fluctuations.

Let us now consider a deterministic growth process where the rate of growth increases faster than the population size. To a rate of growth proportional to $x^2(t)$ there corresponds the differential equation

$$\frac{dx(t)}{dt} = \lambda x^2(t) \tag{4.5}$$

whose solution is

$$x(t) = \frac{i}{1 - \lambda i t}. \tag{4.6}$$

Note that $x(t)$ increases over all bounds as $t \to 1/\lambda i$. In other words, the assumption that the rate of growth increases as the square of the population size implies an infinite growth within a finite time interval. Similarly, if in (3.4) the λ_n increase too fast, there is a finite probability that infinitely many changes take place in a finite time interval. A precise answer about the conditions when such a divergent growth occurs is given by the

Theorem. *In order that* (4.1) *may hold for all t it is necessary and sufficient that the series*

$$\sum \frac{1}{\lambda_n} \tag{4.7}$$

diverge.

Proof. Letting

$$S_k(t) = P_0(t) + \ldots + P_k(t), \tag{4.8}$$

we get from (3.2)

$$S'_k(t) = -\lambda_k P_k(t) \tag{4.9}$$

and hence for $k \geq i$

$$1 - S_k(t) = \lambda_k \int_0^t P_k(\tau)\, d\tau. \tag{4.10}$$

Since all terms in (4.8) are non-negative, the sequence $S_k(t)$—for fixed t—can only increase with k, and therefore the right side in (4.10) decreases monotonically with k. Call its limit $\mu(t)$. Then for $k \geq i$

$$\lambda_k \int_0^t P_k(\tau)\, d\tau \geq \mu(t) \tag{4.11}$$

and hence

$$\int_0^t S_n(\tau)\, d\tau \geq \mu(t) \left(\frac{1}{\lambda_i} + \frac{1}{\lambda_{i+1}} + \ldots + \frac{1}{\lambda_n}\right). \tag{4.12}$$

Because of (4.10) we have $S_n(t) \leq 1$, so that the left side in (4.12) is at most t. If the series (4.7) diverges, the second factor on the right in (4.12) tends to infinity, and the inequality can hold only if $\mu(t) = 0$ for all t. In this case the right side in (4.10) tends to zero as $k \to \infty$,

and therefore $S_n(t) \to 1$, so that (4.1) holds. Conversely,[9] integrating (4.8) and using (4.10) we see that the left side of (4.12) is less than $\lambda_0^{-1} + \lambda_1^{-1} + \ldots + \lambda_n^{-1}$. If the series (4.7) converges, this expression is bounded and hence it is impossible that $S_n(t) \to 1$ for all t.

5. THE BIRTH AND DEATH PROCESS

The pure birth process of section 3 provides a satisfactory description of radioactive transmutations, but it cannot serve as a realistic model for changes in the size of populations whose members can die (or drop out). This suggests generalizing the model by permitting transitions from the state E_n not only to the next higher state E_{n+1} but also to the next lower state E_{n-1}. (More general processes will be defined in section 9.) Accordingly we start from the following

Postulates. *The system changes only through transitions from states to their next neighbors (from E_n to E_{n+1} or E_{n-1} if $n \geq 1$, but from E_0 to E_1 only). If at any time t the system is in state E_n, the probability that during $(t, t+h)$ the transition $E_n \to E_{n+1}$ occurs equals $\lambda_n h + o(h)$, and the probability of $E_n \to E_{n-1}$ (if $n \geq 1$) equals $\mu_n h + o(h)$. The probability that during $(t, t+h)$ more than one change occurs is $o(h)$.*

It is easy to adapt the method of section 2 to derive differential equations for the probabilities $P_n(t)$ of finding the system at time t in state E_n. To calculate $P_n(t+h)$, note that at time $t+h$ the system can be in state E_n only if one of the following conditions is satisfied: (1) At time t the system is in E_n and during $(t, t+h)$ no change occurs; (2) at time t the system is in E_{n-1} and a transition to E_n occurs; (3) at time t the system is in E_{n+1} and a transition to E_n occurs; (4) during $(t, t+h)$ two or more transitions occur. By assumption, the probability of the last event is $o(h)$. The first three contingencies are mutually exclusive, so that their probabilities add. Therefore

$$\begin{aligned} P_n(t+h) = P_n(t)\{1 - \lambda_n h - \mu_n h\} &+ \\ + \lambda_{n-1} h P_{n-1}(t) &+ \mu_{n+1} h P_{n+1}(t) + o(h). \end{aligned} \tag{5.1}$$

Transposing the term $P_n(t)$ and dividing the equation by h, we get on the left the difference ratio of $P_n(t)$. Letting $h \to 0$, we get

$$P'_n(t) = -(\lambda_n + \mu_n) P_n(t) + \lambda_{n-1} P_{n-1}(t) + \mu_{n+1} P_{n+1}(t). \tag{5.2}$$

[9] By a regrettable oversight the following three lines were missing in the first printing of the first edition and part of the preceding argument was repeated instead. The error was corrected after a few months. (The present discussion is continued in section 10.)