relation $\text{Var}(\mathbf{X}_{n+1}) = \mu^2 \text{Var}(\mathbf{X}_n) + \mu^n\sigma^2$. Conclude from either form that $\text{Var}(\mathbf{X}_n) = \sigma^2(\mu^{2n-2} + \mu^{2n-3} + \cdots + \mu^{n-1})$.

8. *Continuation.* If $n > m$ show that $\mathbf{E}(\mathbf{X}_n\mathbf{X}_m) = \mu^{n-m}\mathbf{E}(\mathbf{X}_m^2)$.

9. *Continuation.* Show that the bivariate generating function of $(\mathbf{X}_m, \mathbf{X}_n)$ is $P_m(s_1P_{n-m}(s_2))$. Use this to verify the assertion in problem 8.

10. *Branching processes with two types of individuals.* Assume that each individual can have descendants of either kind; the numbers of descendants of the two types are regulated by two bivariate generating functions $P_1(s_1, s_2)$ and $P_2(s_1, s_2)$. We have now two extinction probabilities x, y depending on the type of the ancestor. Show that the pair (x, y) satisfies the equations

$$x = P_1(x, y), \qquad y = P_2(x, y). \tag{6.1}$$

Prove that these equations have at most one solution $0 \le x \le 1$, $0 \le y \le 1$ different from $(1, 1)$. The solution $(1, 1)$ is unique if, and only if, $\mu_{11} \le 1$, $\mu_{22} \le 1$ and $(1 - \mu_{11})({}_1 - \mu_{22}) \ge \mu_{12}\mu_{21}$ where $\mu_{ij} = \dfrac{\partial P_i(1, 1)}{\partial s_j}$.

CHAPTER XIII

Recurrent Events. Renewal Theory

1. INFORMAL PREPARATIONS AND EXAMPLES

We shall be concerned with certain repetitive, or recurrent, patterns connected with repeated trials. Roughly speaking, a pattern $\mathcal{E}$ qualifies for the following theory if after each occurrence of $\mathcal{E}$ the trials start from scratch in the sense that the trials following an occurrence of $\mathcal{E}$ form a replica of the whole experiment. The waiting times between successive occurrences of $\mathcal{E}$ are mutually independent random variables having the same distribution.

The simplest special case arises when $\mathcal{E}$ stands as abbreviation for "a success occurs" in a sequence of Bernoulli trials. The waiting time up to the first success has a geometric distribution; when the first success occurs, the trials start anew, and the number of trials between the rth and the $(r+1)$st success has the same geometric distribution. The waiting time up to the rth success is the sum of r independent variables [example IX,(3.*c*)]. This situation prevails also when $\mathcal{E}$ stands for "a success followed by failure": The occurrence of the pattern *SF* reestablishes the initial situation, and the waiting time for the next occurrence of *SF* is independent of the preceding trials. By contrast, suppose that people are sampled one by one and let $\mathcal{E}$ stand for "two people in the sample have the same birthday." This $\mathcal{E}$ is not repetitive because after its first realization $\mathcal{E}$ persists forever. If we change the definition to "$\mathcal{E}$ occurs whenever the birthday of the newly added person is already present in the sample," then $\mathcal{E}$ can occur any number of times, but after an occurrence of $\mathcal{E}$ the process does *not* start from scratch. This is so because the increasing sample size makes duplications of birthdays more likely; a long waiting time for the first double birthday promises therefore a shorter waiting time for the second duplication, and so the consecutive waiting times are neither independent nor subject to a common distribution.

The importance of the theory of recurrent patterns is due to the fact that such patterns occur frequently in connection with various sequences of variables (stochastic processes). The laws governing a sequence of random variables may be so intricate as to preclude a complete analysis but the existence of a repetitive pattern makes it always possible to discuss essential features of the sequence, to prove the existence of certain limits, etc. This approach contributes greatly to the simplification and unification of many theories.

We proceed to review a few typical examples, some of which are of intrinsic interest. The first examples refer to the familiar Bernoulli trials, but the last three involve more complicated schemes. In their description we use terms such as "server" and "customer," but in each case we give a mathematical definition of a sequence of random variables which is complete in the sense that it uniquely determines the probabilities of all possible events. In practice, not even the basic probabilities can be calculated explicitly, but it will turn out that the theory of repetitive patterns nevertheless leads to significant results.

Examples. (*a*) *Return to equilibrium.* In a sequence of Bernoulli trials let $\mathcal{E}$ stand as abbreviation for "the accumulated numbers of successes and failures are equal." As we have done before, we describe the trials in terms of mutually independent random variables $\mathbf{X}_1, \mathbf{X}_2, \ldots$ assuming the values 1 and -1 with probabilities p and q, respectively. As usual, we put

$$(1.1)\qquad \mathbf{S}_0 = 0, \qquad \mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n.$$

Then $\mathbf{S}_n$ is the accumulated excess of heads over tails, and $\mathcal{E}$ *occurs if, and only if,* $\mathbf{S}_n = 0$. It goes without saying that the occurrence of this event reestablishes the initial situation in the sense that the subsequent partial sums $\mathbf{S}_{n+1}, \mathbf{S}_{n+2}, \ldots$ form a probabilistic replica of the whole sequence $\mathbf{S}_1, \mathbf{S}_2, \ldots$. [Continued in example (4.*b*).]

(*b*) *Return to equilibrium through negative values.* We modify the last example by stipulating that $\mathcal{E}$ occurs at the *n*th trial if

$$(1.2)\qquad \mathbf{S}_n = 0, \quad but \quad \mathbf{S}_1 < 0, \ldots, \mathbf{S}_{n-1} < 0.$$

Again, it is clear that the occurrence of $\mathcal{E}$ implies that we start from scratch. [Continued in example (4.*c*).]

(*c*) Another variant of example (*a*) is the event $\mathcal{E}$ that the accumulated number of successes equals λ times the accumulated number of failures (where $\lambda > 0$ is an arbitrary, but fixed, number). If $\mathcal{E}$ occurs at the *n*th trial, it occurs again at the $(n+m)$th trial only if among the trials number $n+1, \ldots, n+m$ there occur exactly λ times as many successes as failures. The waiting times between successive occurrences of $\mathcal{E}$ are therefore independent and identically distributed. As a special case consider the event that $6n$ throws of a perfect die yield exactly n aces. (Continued in problems 4–5.)

(*d*) *Ladder variables.* Adhering to the notations of example (*a*) we define a new repetitive pattern $\mathcal{E}$ by stipulating that $\mathcal{E}$ *occurs at the nth trial if* $\mathbf{S}_n$ *exceeds all preceding sums*, that is, if

$$(1.3)\qquad \mathbf{S}_n > 0, \qquad \mathbf{S}_n > \mathbf{S}_1, \ldots, \mathbf{S}_n > \mathbf{S}_{n-1}.$$

If $\mathcal{E}$ occurs at the *n*th trial the process starts from scratch in the following sense. Assuming (1.3) to hold, $\mathcal{E}$ occurs at the $(n+m)$th trial if, and only if,

$$(1.4)\qquad \mathbf{S}_{n+m} > \mathbf{S}_n, \ldots, \mathbf{S}_{n+m} > \mathbf{S}_{n+m-1}.$$

But the differences $\mathbf{S}_{n+k} - \mathbf{S}_n$ are simply the partial sums of the residual sequence $\mathbf{X}_{n+1}, \mathbf{X}_{n+2}, \ldots$ and so the reoccurrence of $\mathcal{E}$ is defined in terms of this residual sequence exactly as $\mathcal{E}$ is defined for the whole sequence. In other words, for the study of $\mathcal{E}$ the whole past becomes irrelevant every time $\mathcal{E}$ occurs. [Continued in example (4.*d*).]

(*e*) *Success runs in Bernoulli trials.* In the preceding examples the definition of $\mathcal{E}$ was straightforward, but we turn now to a situation in which a judicious definition is necessary to make the theory of recurrent patterns applicable. In the classical literature a "success run of length r" meant an uninterrupted sequence of either exactly r, or of at least r, successes. Neither convention leads to a recurrent pattern. Indeed, if exactly r successes are required, then a success at the $(n+1)$st trial may undo the run completed at the *n*th trial. On the other hand, if at least r successes are required, then every run may be prolonged indefinitely and it is clear that the occurrence of a run does not reestablish the initial situation. The classical theory of runs was rather messy, and a more systematic approach is possible by defining a run of length r in such a way that it becomes a recurrent pattern. A *first* run of length r is uniquely defined, and we now agree to start counting from scratch every time a run occurs. With this convention the sequence $SSS\,|\,SFSSS\,|\,SSS\,|\,F$ contains three success runs of length three (occurring at trials number 3, 8, and 11). It contains five runs of length two (trials number 2, 4, 7, 9, 11). The formal definition is as follows: *A sequence of n letters S and F contains as many S-runs of length r as there are non-overlapping uninterrupted blocks containing exactly r letters S each.* With this convention we say that $\mathcal{E}$ occurs at the *n*th trial if a new run of length r is added to the sequence. This defines a recurrent pattern and greatly simplifies the theory without affecting its basic features. (Continued in section 7.)

(*f*) *Continuation: Related patterns.* It is obvious that the considerations of the preceding example apply to more general patterns, such as the occurrence of the succession *SFSF*. More interesting is that no limitation to a fixed pattern is necessary. Thus the occurrence of "*two successes and three failures*" defines a repetitive pattern, and the same is true of "either a success run of length r or a failure run of length ρ." (Continued in section 8.)

(*g*) *Geiger counters.* Counters of the type used for cosmic rays and α-particles may be described by the following simplified model.[1] Bernoulli trials are performed at a uniform rate. A counter is meant to register successes, but the mechanism is locked for exactly $r-1$ trials following each registration. In other words, a success at the nth trial is registered if, and only if, no registration has occurred in the preceding $r-1$ trials. The counter is then locked at the conclusion of trials number $n, \ldots, n+r-1$, and is freed at the conclusion of the $(n+r)$th trial provided this trial results in failure. The output of the counter represents dependent trials. Each registration has an aftereffect, but, whenever the counter is free (not locked) the situation is exactly the same, and the trials start from scratch. Letting $\mathcal{E}$ stand for "at the conclusion of the trial the counter is free," we have a typical recurrent pattern. [Continued in example (4.*e*).]

(*h*) *The simplest queuing process* is defined in terms of a sequence of Bernoulli trials and a sequence of random variables $\mathbf{X}_1, \mathbf{X}_2, \ldots$ assuming only positive integral values. The $\mathbf{X}_k$ have a common distribution $\{\beta_k\}$ and are independent of each other and of the Bernoulli trials. We interpret success at the nth trial as the arrival at epoch[2] n of a customer at a server (or a call at a telephone trunk line). The variable $\mathbf{X}_n$ represents the service time of the nth customer arriving at the server. At any epoch the server is either "free" or "busy" and the process proceeds according to the following rules. Initially (at epoch 0) the server is free. A customer arriving when the counter is free is served immediately, but following his arrival the server is busy for the duration of the service time. Customers arriving when the server is busy form a waiting line (queue). The server serves customers without interruption as long as there is a demand.

These rules determine the process uniquely, for given a sample sequence $(S, F, S, S, S, F, F, \ldots)$ for the arrival process and a sample sequence $(3, 1, 17, 2, \ldots)$ for the successive service times, it is not difficult to find

[1] This is the discrete analogue of the so-called counters of type I. Type II is described in problem 8.

[2] We use the term *epoch* to denote points on the time axis. Terms such as waiting time will refer to durations. This practice was introduced by J. Riordan because in queuing theory the several meanings of words like time, moment, etc., are apt to cause confusion.

the size of the queue at any epoch, and the waiting time of the nth customer. In principle, therefore, we should be able to calculate all pertinent probabilities, but it is not easy to find practicable methods. Now it is clear that every time the server is free the situation is exactly the same as it is at epoch 0. In our terminology therefore the contingency "the server is free" constitutes a recurrent pattern. We shall see that the very existence of such a recurrent pattern has important consequences; for example, it implies that the probability distributions for the size of the queue at epoch n, for the waiting time of the nth customer, and for similar random variables tend to definite limits when $n \to \infty$ (theorem 5.2). In other words, the existence of a recurrent pattern enables us to prove the existence of a steady state and to analyze its dominant features.

(*i*) *Servicing of machines.* The scope of the method of recurrent patterns may be illustrated by a variant of the preceding example in which the arrivals are no longer regulated by Bernoulli trials. To fix ideas, let us interpret the "customers" as identical machines subject to occasional breakdowns, and the "server" as a repairman. We adhere to the same conventions concerning servicing and the formation of queues, but introduce a new chance mechanism for the "arrivals," that is, for the breakdowns. Suppose there are N machines in all, and consider two extreme cases.

(*a*) Suppose first that as long as a machine is in working condition it has a fixed probability p to break down at the next epoch; when it breaks down it is replaced by an identical new machine, and the serving time is interpreted as the time required for the installation of a new machine. We treat the machines as independent, and the breakdowns are regulated by N independent sequences of Bernoulli trials. Note that the more machines are in the queue, the fewer machines are in working condition, and hence the length of the queue at any epoch influences the probability of new arrivals (or service calls). This is in marked contrast to the preceding example, but the contingency "server is idle" constitutes nevertheless a recurrent pattern because we are confronted with precisely the same situation whenever it occurs.

(*b*) Suppose now that every repair has an aftereffect in that it increases the probabilities of further breakdowns. This implies that the machines deteriorate steadily and so once a machine breaks down it is impossible that the favorable initial situation should be repeated. In this case there is no recurrent pattern to help the analysis. ►

2. DEFINITIONS

We consider a sequence of repeated trials with possible outcomes E_j $(j = 1, 2, \ldots)$. They need not be independent (applications to Markov chains being of special interest). As usual, we suppose that it is in principle possible to continue the trials indefinitely, the probabilities $\mathbf{P}\{E_{j_1}, E_{j_2}, \ldots, E_{j_n}\}$ being defined consistently for all finite sequences. Let $\mathcal{E}$ be an attribute of finite sequences; that is, we suppose that it is uniquely determined whether a sequence $(E_{j_1}, \ldots, E_{j_n})$ has, or has not, the characteristic $\mathcal{E}$. We agree that the expression "$\mathcal{E}$ occurs at the nth place in the (finite or infinite) sequence $E_{j_1}, E_{j_2}, \ldots$" is an abbreviation for "The subsequence $E_{j_1}, E_{j_2}, \ldots, E_{j_n}$ has the attribute $\mathcal{E}$." This convention implies that the occurrence of $\mathcal{E}$ at the nth trial depends soley on the outcome of the first n trials. It is also understood that *when speaking of a "recurrent event $\mathcal{E}$," we are really referring to a class of events* defined by the property that $\mathcal{E}$ occurs. Clearly $\mathcal{E}$ itself is a label rather than an event. We are here abusing the language in the same way as is generally accepted in terms such as "a two-dimensional problem"; the problem itself is dimensionless.

Definition 1. *The attribute $\mathcal{E}$ defines a recurrent event if:*

(*a*) *In order that $\mathcal{E}$ occurs at the nth and the $(n+m)$th place of the sequence $(E_{j_1}, E_{j_2}, \ldots, E_{j_{n+m}})$ it is necessary and sufficient that $\mathcal{E}$ occurs at the last place in each of the two subsequences $(E_{j_1}, E_{j_2}, \ldots, E_{j_n})$ and $(E_{j_{n+1}}, E_{j_{n+2}}, \ldots, E_{j_{n+m}})$.*

(*b*) *If $\mathcal{E}$ occurs at the nth place then identically*

$$\mathbf{P}\{E_{j_1}, \ldots, E_{j_{n+m}}\} = \mathbf{P}\{E_{j_1}, \ldots, E_{j_n}\}\,\mathbf{P}\{E_{j_{n+1}}, \ldots, E_{j_{n+m}}\}.$$

It has now an obvious meaning to say that $\mathcal{E}$ occurs in the sequence $(E_{j_1}, E_{j_2}, \ldots)$ *for the first time* at the nth place, etc. It is also clear that with each recurrent event $\mathcal{E}$ there are associated the two sequences of numbers defined for $n = 1, 2, \ldots$ as follows

$$(2.1)\qquad \begin{aligned} u_n &= \mathbf{P}\{\mathcal{E} \text{ occurs at the } n\text{th trial}\},\\ f_n &= \mathbf{P}\{\mathcal{E} \text{ occurs for the first time at the } n\text{th trial}\}. \end{aligned}$$

It will be convenient to define

$$(2.2)\qquad f_0 = 0, \qquad u_0 = 1,$$

and to introduce the generating functions

$$(2.3)\qquad F(s) = \sum_{k=1}^{\infty} f_k s^k, \qquad U(s) = \sum_{k=0}^{\infty} u_k s^k.$$

Observe that $\{u_k\}$ is not a probability distribution; in fact, in representative cases we shall have $\sum u_k = \infty$. However, the events "$\mathcal{E}$ occurs for the first time at the nth trial" are mutually exclusive, and therefore

$$(2.4)\qquad f = F(1) = \sum_{n=1}^{\infty} f_n \leq 1.$$

It is clear that $1 - f$ should be interpreted as *the probaiblity that $\mathcal{E}$ does not occur in an indefinitely prolonged sequence of trials.* If $f = 1$ we may introduce a random variable $\mathbf{T}$ with distribution

$$(2.5)\qquad \mathbf{P}\{\mathbf{T} = n\} = f_n.$$

We shall use the same notation (2.5) even if $f < 1$. Then $\mathbf{T}$ *is an improper, or defective, random variable, which with probability $1 - f$ does not assume a numerical value.* (For our purposes we could assign to $\mathbf{T}$ the symbol ∞, and it should be clear that no new rules are required.)

The *waiting time for* $\mathcal{E}$, that is, the number of trials up to and including the first occurrence of $\mathcal{E}$, is a random variable with the distribution (2.5); however, this random variable is really defined only in the space of infinite sequences $(E_{j_1}, E_{j_2}, \ldots)$.

By the definition of recurrent events the probability that $\mathcal{E}$ occurs for the first time at trial number k and for the *second* time at the nth trial equals $f_k f_{n-k}$. Therefore the probability $f_n^{(2)}$ that $\mathcal{E}$ occurs for the second time at the nth trial equals

$$(2.6)\qquad f_n^{(2)} = f_1 f_{n-1} + f_2 f_{n-2} + \cdots + f_{n-1} f_1.$$

The right side is the convolution of $\{f_n\}$ with itself and therefore $\{f_n^{(2)}\}$ represents the probability distribution of the sum of two independent random variables each having the distribution (2.5). More generally, if $f_n^{(r)}$ is the probability that the rth occurrence of $\mathcal{E}$ takes place at the nth trial we have

$$(2.7)\qquad f_n^{(r)} = f_1 f_{n-1}^{(r-1)} + f_2 f_{n-2}^{(r-1)} + \cdots + f_{n-1} f_1^{(r-1)}.$$

This simple fact is expressed in the

Theorem. *Let $f_n^{(r)}$ be the probability that the rth occurrence of $\mathcal{E}$ takes place at the nth trial. Then $\{f_n^{(r)}\}$ is the probability distribution of the sum*

$$(2.8)\qquad \mathbf{T}^{(r)} = \mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_r$$

of r independent random variables $\mathbf{T}_1, \ldots, \mathbf{T}_r$ each having the distribution (2.5). *In other words: For fixed r the sequence $\{f_n^{(r)}\}$ has the generating function $F^r(s)$.*

It follows in particular that

$$\sum_{n=1}^{\infty} f_n^{(r)} = F^r(1) = f^r. \tag{2.9}$$

In words: *the probability that* $\mathcal{E}$ *occurs at least* r *times equals* f^r (a fact which could have been anticipated). We now introduce

Definition 2. *A recurrent event* $\mathcal{E}$ *will be called persistent*[3] *if* $f = 1$ *and transient if* $f < 1$.

For a transient $\mathcal{E}$ the probability f^r that it occurs at least r times tends to zero, whereas for a persistent $\mathcal{E}$ this probability remains unity. This can be described by saying *with probability one a persistent* $\mathcal{E}$ *is bound to occur infinitely often whereas a transient* $\mathcal{E}$ *occurs only a finite number of times.* (This statement not only is a description but is formally correct if interpreted in the sample space of infinite sequences $E_{j_1}, E_{j_2}, \ldots$.)

We require one more definition. In Bernoulli trials a return to equilibrium [example (1.*a*)] can occur only at an *even*-numbered trial. In this case $f_{2n+1} = u_{2n+1} = 0$, and the generating functions $F(s)$ and $U(s)$ are power series in s^2 rather than s. Similarly, in example (1.*c*) if λ is an integer, $\mathcal{E}$ can occur at the nth trial only if n is a multiple of $\lambda + 1$. We express this by saying that $\mathcal{E}$ is periodic. In essence periodic recurrent events differ only notationally from non-periodic ones, but every theorem requires a special mention of the exceptional periodic case. In other words, periodic recurrent events are a great nuisance without redeeming features of interest.

Definition 3. *The recurrent event* $\mathcal{E}$ *is called periodic if there exists an integer* $\lambda > 1$ *such that* $\mathcal{E}$ *can occur only at trials number* $\lambda, 2\lambda, 3\lambda, \ldots$ (*i.e.,* $u_n = 0$ *whenever* n *is not divisible by* λ). *The greatest* λ *with this property is called the period of* $\mathcal{E}$.

In conclusion let us remark that in the sample space of infinite sequences $E_{j_1}, E_{j_2}, \ldots$ the number of trials between the $(r-1)$st and the rth occurrence of $\mathcal{E}$ is a well-defined random variable (possibly a defective one), having the probability distribution of our $\mathbf{T}_r$. In other words, our variables $\mathbf{T}_r$ really stand for the *waiting times between the successive occurrences of* $\mathcal{E}$ (*the recurrence times*). We have defined the $\mathbf{T}_r$ analytically in order not to refer to sample spaces beyond the scope of this volume, but it is hoped that the probabilistic background appears in all its intuitive simplicity. The notion of recurrent events is designed to

reduce a fairly general situation to sums of independent random variables. Conversely, *an arbitrary probability distribution* $\{f_n\}$, $n = 1, 2, \ldots$ *may be used to define a recurrent event.* We prove this assertion by the

Example. *Self-renewing aggregates.* Consider an electric bulb, fuse, or other piece of equipment with a finite life span. As soon as the piece fails, it is replaced by a new piece of like kind, which in due time is replaced by a third piece, and so on. We assume that the life span is a random variable which ranges only over multiples of a unit time interval (year, day, or second). Each time unit then represents a trial with possible outcomes "replacement" and "no replacement." The successive replacements may be treated as recurrent events. If f_n is the probability that a new piece will serve for exactly n time units, then $\{f_n\}$ is the distribution of the recurrence times. When it is certain that the life span is finite, then $\sum f_n = 1$ and the recurrent event is persistent. Usually it is known that the life span cannot exceed a fixed number m, in which case the generating function $F(s)$ is a polynomial of a degree not exceeding m. In applications we desire the probability u_n that a replacement takes place at time n. This u_n may be calculated from (3.1). Here we have a class of recurrent events defined solely in terms of an arbitrary distribution $\{f_n\}$. The case $f < 1$ is not excluded, $1 - f$ being the probability of an eternal life of our piece of equipment. ▶

3. THE BASIC RELATIONS

We adhere to the notations (2.1)–(2.4) and propose to investigate the connection between the $\{f_n\}$ and the $\{u_n\}$. The probability that $\mathcal{E}$ occurs for the first time at trial number ν and then again at a later trial $n > \nu$ is, by definition, $f_\nu u_{n-\nu}$. The probability that $\mathcal{E}$ occurs at the nth trial for the first time is $f_n = f_n u_0$. Since these cases are mutually exclusive we have

$$u_n = f_1 u_{n-1} + f_2 u_{n-2} + \cdots + f_n u_0, \qquad n \geq 1. \tag{3.1}$$

At the right we recognize the convolution $\{f_k\} * \{u_k\}$ with the generating function $F(s)\, U(s)$. At the left we find the sequence $\{u_n\}$ with the term u_0 missing, so that its generating function is $U(s) - 1$. Thus $U(s) - 1 = F(s)\, U(s)$, and we have proved

Theorem 1. *The generating functions of* $\{u_n\}$ *and* $\{f_n\}$ *are related by*

$$U(s) = \frac{1}{1 - F(s)}. \tag{3.2}$$

[3] In the first edition the terms certain and uncertain were used, but the present terminology is preferable in applications to Markov chains.

Note. The right side in (3.2) can be expanded into a geometric series $\sum F^r(s)$ converging for $|s| < 1$. The coefficient $f_n^{(r)}$ of s^n in $F^r(s)$ being the probability that the rth occurrence of $\mathcal{E}$ takes place at the nth trial, (3.2) is equivalent to

$$u_n = f_n^{(1)} + f_n^{(2)} + \cdots ; \tag{3.3}$$

this expresses the obvious fact that if $\mathcal{E}$ occurs at the nth trial, it has previously occurred $0, 1, 2, \ldots, n-1$ times. (Clearly $f_n^{(r)} = 0$ for $r > n$.)

Theorem 2. *For $\mathcal{E}$ to be transient, it is necessary and sufficient that*

$$u = \sum_{j=0}^{\infty} u_j \tag{3.4}$$

is finite. In this case the probability f that $\mathcal{E}$ ever occurs is given by

$$f = \frac{u-1}{u}. \tag{3.5}$$

Note. We can interpret u_j as the expectation of a random variable which equals 1 or 0 according to whether $\mathcal{E}$ does or does not occur at the jth trial. Hence $u_1 + u_2 + \cdots + u_n$ *is the expected number of occurrences of $\mathcal{E}$ in n trials,* and $u - 1$ can be interpreted as the expected number of occurrences of $\mathcal{E}$ in infinitely many trials.

Proof. The coefficients u_k being non-negative, it is clear that $U(s)$ increases monotonically as $s \to 1$ and that for each N

$$\sum_{n=0}^{N} u_n \leq \lim_{s \to 1} U(s) \leq \sum_{n=0}^{\infty} u_n = u.$$

Since $U(s) \to (1-f)^{-1}$ when $f < 1$ and $U(s) \to \infty$ when $f = 1$, the theorem follows. ►

The next theorem is of particular importance.[4] The proof is of an

elementary nature, but since it does not contribute to a probabilistic understanding we defer it to the end of the chapter.

Theorem 3. *Let $\mathcal{E}$ be persistent and not periodic and denote by μ the mean of the recurrence times* $\mathbf{T}_\nu$, *that is,*

$$\mu = \sum j f_j = F'(1) \tag{3.6}$$

(possibly $\mu = \infty$). Then

$$u_n \to \mu^{-1} \tag{3.7}$$

as $n \to \infty$ ($u_n \to 0$ if the mean recurrence time is infinite).

The restriction to non-periodic $\mathcal{E}$ is easily removed. In fact, when $\mathcal{E}$ has period λ the series $\sum f_n s^n$ contains only powers of s^λ. Let us call a power series honest if this is not the case for any integer $\lambda > 1$. Theorem 3 may then be restated to the effect, that *if F is an honest probability generating function and U is defined by* (3.2), *then* $u_n \to 1/F'(1)$. Now if $\mathcal{E}$ has period λ then $F(s^{1/\lambda})$ is an honest probability generating function, and hence the coefficients of $U(s^{1/\lambda})$ converge to $\lambda/F'(1)$. We have thus

Theorem 4. *If $\mathcal{E}$ is persistent and has period λ then*

$$u_{n\lambda} \to \lambda/\mu \tag{3.8}$$

while $u_k = 0$ for every k not divisible by λ.

4. EXAMPLES

(*a*) *Successes in Bernoulli trials.* For a trite example let $\mathcal{E}$ stand for "success" in a sequence of Bernoulli trials. Then $u_n = p$ for $n \geq 1$, whence

$$U(s) = \frac{1-qs}{1-s} \quad \textit{and therefore} \quad F(s) = \frac{ps}{1-qs} \tag{4.1}$$

by virtue of (3.2). In this special case theorem 2 merely confirms the obvious fact that the waiting times between consecutive successes have a geometric distribution with expectation $1/p$.

(*b*) *Returns to equilibrium* [*example* (1.*a*)]. At the kth trial the accumulated numbers of heads and tails can be equal only if $k = 2n$ is even, and in this case the probability of an equalization equals

$$u_{2n} = \binom{2n}{n} p^n q^n = \binom{-\frac{1}{2}}{n} (-4pq)^n. \tag{4.2}$$

[4] Special cases are easily proved (see problem 1) and were known for a long time. A huge literature tried to improve on the conditions, but it was generally believed that some restrictions were necessary. In full generality theorem 3 was proved by P. Erdös, W. Feller, and H. Pollard, A theorem on power series, Bull. Amer. Math. Soc. vol. 55 (1949), pp. 201–204. After the appearance of the first edition it was observed by K. L. Chung that the theorem could be derived from Kolmogorov's results about the asymptotic behavior of Markov chains. Many prominent mathematicians proved various extensions of the theorem to different classes of probability distributions. These investigations contributed to the methodology of modern probability theory. Eventually it turned out that an analogue to theorem 3 holds for arbitrary probability distributions. For an elementary (if not simple) proof see XI,9 of volume 2.

From the binomial expansion II, (8.7) it follows therefore that

$$U(s) = \frac{1}{\sqrt{1-4pqs^2}} \tag{4.3}$$

and hence from (3.2)

$$F(s) = 1 - \sqrt{1-4pqs^2}. \tag{4.4}$$

A second application of the binomial expansion leads to an explicit expression for f_{2n}. (Explicit expressions for u_{2n} and f_{2n} when $p = \frac{1}{2}$ were derived by combinatorial methods in III,2-3; the generating functions U and F were found by other methods in XI,3. It will be noticed that only the present method requires no artifice.)

For $s = 1$ the square root in (4.4) equals $|p - q|$ and so

$$f = 1 - |p - q|. \tag{4.5}$$

Thus *returns to equilibrium represent a recurrent event with period* 2 *which is transient when* $p \neq q$, *and persistent in the symmetric case* $p = q$. The probability of at least r returns to equilibrium equals f^r.

When $p = q = \frac{1}{2}$ the waiting time for the first return to equilibrium is a proper random variable, but $F'(1) = \infty$ and so *the mean recurrence time* μ *is infinite*. (This follows also from theorem 4 and the fact that $u_n \to 0$.) The fact that the mean recurrence time is infinite implies that the chance fluctuations in an individual prolonged coin-tossing game are far removed from the familiar pattern governed by the normal distribution. The rather paradoxical true nature of these fluctuations was discussed in chapter III.

(*c*) *Return to equilibrium through negative values.* In example (1.*b*) the return to equilibrium was subject to the restriction that no preceding partial sum $\mathbf{S}_j$ was positive. The distribution of the recurrence times for this recurrent event is defined by

$$f_{2n}^- = \mathbf{P}\{\mathbf{S}_{2n} = 0, \mathbf{S}_1 < 0, \ldots, \mathbf{S}_{2n-1} < 0\} \tag{4.6}$$

and, of course, $f_{2n-1}^- = 0$. It does not seem possible to find these probabilities by a direct argument, but they follow easily from the preceding example. Indeed, a sample sequence $(\mathbf{X}_1, \ldots, \mathbf{X}_{2n})$ satisfying the condition in (4.6) contains n plus ones and n minus ones, and hence it has the same probability as $(-\mathbf{X}_1, \ldots, -\mathbf{X}_{2n})$. Now a *first* return to equilibrium occurs either through positive or through negative values, and we conclude that these two contingencies have the same probability. Thus $f_{2n}^- = \frac{1}{2}f_{2n}$ where $\{f_n\}$ is the distribution for the returns to equilibrium found in the preceding example. The generating function for our recurrence times is therefore given by

$$F^-(s) = \tfrac{1}{2} - \tfrac{1}{2}\sqrt{1-4pqs^2}, \tag{4.7}$$

and hence

$$U^-(s) = \frac{2}{1+\sqrt{1-4pqs^2}} = \frac{1-\sqrt{1-4pqs^2}}{2pqs^2}. \tag{4.8}$$

The event $\mathcal{E}$ is transient, the probability that it ever occurs being $\frac{1}{2} - \frac{1}{2}|p - q|$.

(*d*) *Ladder variables.* The first positive partial sum can occur at the kth trial only if $k = 2n + 1$ is odd. For the corresponding probabilities we write

$$\phi_{2n+1} = \mathbf{P}\{\mathbf{S}_1 < 0, \ldots, \mathbf{S}_{2n} = 0, \mathbf{S}_{2n+1} = 1\}. \tag{4.9}$$

Thus $\{\phi_k\}$ is the distribution of the recurrent event of example (1.*d*). Now the condition in (4.9) requires that $\mathbf{X}_{2n+1} = +1$, and that the recurrent event of the preceding example occurs at the $2n$th trial. It follows that $\phi_{2n+1} = p \cdot u_{2n}^-$. With obvious notations therefore

$$\Phi(s) = psU^-(s) = \frac{1-\sqrt{1-4pqs^2}}{2qs}. \tag{4.10}$$

This is the generating function for the first-passage times found in XI,(3.6). An explicit expression for ϕ_{2n+1} follows from (4.10) using the binomial expansion II,(8.7). This expression for ϕ_{2n+1} agrees with that found by combinatorial methods in theorem 2 of III,7.

(*e*) *Geiger counters.* In example (1.*g*) the counter remains free if no registration takes place at epoch 1. Otherwise it becomes locked and is freed again at epoch $r + 1$ if no particle arrives at that epoch; the counter is freed at epoch $2r + 1$ if a particle appears at epoch $r + 1$, but none at epoch $2r + 1$, and so on. The generating function of the recurrence times is therefore given by

$$qs + qps^{r+1} + qp^2s^{2r+1} + \cdots = \frac{qs}{1-ps^r}. \tag{4.11}$$

(See also problems 7–9.)

(*f*) *The simplest queuing problem* [example (1.*h*)]. Here the server remains free if no customer arrives at epoch 1. If a customer arrives there follows a so-called "busy period" which terminates at the epoch when the counter first becomes free. The generating function $\rho(s)$ for the busy period was derived in example XII,(5.*c*) using the methods of branching processes.

It follows that in the present case the generating function of the recurrence times is given by $qs + ps\rho(s)$, in agreement with XII,(5.7).

(*g*) *Ties in multiple coin games.* We conclude with a simple example showing the possibility of certain conclusions without explicit knowledge of the generating functions. Let $r \geq 2$ be an arbitrary integer and consider a sequence of simultaneous independent tosses of r coins. Let $\mathcal{E}$ stand for the recurrent event that *all* r *coins are in the same phase* (that is, the accumulated numbers of heads are the same for all r coins). The probability that this occurs at the nth trial is

$$u_n = 2^{-rn}\left[\binom{n}{0}^r + \binom{n}{1}^r + \cdots + \binom{n}{n}^r\right]. \tag{4.12}$$

On the right we recognize the terms of the binomial distribution with $p = \frac{1}{2}$, and from the normal approximation to the latter we conclude easily that for each fixed r as $n \to \infty$

$$u_n \sim \left(\frac{2}{\pi n}\right)^{\frac{1}{2}r} \sum_j e^{-2rj^2/n} \tag{4.13}$$

(the summation extending over all integers j between $-\frac{1}{2}n$ and $\frac{1}{2}n$). But by the very definition of the integral

$$2\sqrt{\frac{r}{n}} \sum_j e^{-2rj^2/n} \to \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}\,dx = \sqrt{2\pi} \tag{4.14}$$

and hence we conclude that

$$u_n \sim \frac{1}{\sqrt{r}}\left(\frac{2}{\pi n}\right)^{\frac{1}{2}(r-1)}. \tag{4.15}$$

This implies that $\sum u_n$ diverges when $r \leq 3$, but converges when $r \geq 4$. It follows that $\mathcal{E}$ *is persistent when* $r \leq 3$ *but transient if* $r \geq 4$. Since $u_n \to 0$ the mean recurrence time is infinite when $r \leq 3$. (Compare problems 2 and 3.) ▶

5. DELAYED RECURRENT EVENTS. A GENERAL LIMIT THEOREM

We shall now introduce a slight extension of the notion of recurrent events which is so obvious that it could pass without special mention, except that it is convenient to have a term for it and to have the basic equations on record.

Perhaps the best informal description of delayed recurrent events is to say that they refer to trials where we have "missed the beginning and start in the middle." The waiting time up to the *first* occurrence of $\mathcal{E}$ has a distribution $\{b_n\}$ different from the distribution $\{f_n\}$ of the recurrence times between the following occurrences of $\mathcal{E}$. The theory applies without change except that the trials following each occurrence of $\mathcal{E}$ are exact replicas of a fixed sample space which is not identical with the original one.

The situation being so simple, we shall forego formalities and agree to speak of a *delayed recurrent* $\mathcal{E}$ *when the definition of recurrent events applies only if the trials leading up to the first occurrence of* $\mathcal{E}$ *are disregarded; it is understood that the waiting time up to the first appearance of* $\mathcal{E}$ *is a random variable independent of the following recurrence times, although its distribution* $\{b_n\}$ *may be different from the common distribution* $\{f_n\}$ *of the recurrence times.*

We denote by v_n *the probability of the occurrence of* $\mathcal{E}$ *at the nth trial.* To derive an expression for v_n we argue as follows. Suppose that $\mathcal{E}$ occurs at trial number $k < n$. Relative to the subsequent trials $\mathcal{E}$ becomes an ordinary recurrent event and so the (conditional) probability of a renewed occurrence at the nth trial equals u_{n-k}. Now if $\mathcal{E}$ occurs at the nth trial this is either its first occurrence, or else the first occurrence took place at the kth trial for some $k < n$. Summing over all possibilities we get

$$v_n = b_n + b_{n-1}u_1 + b_{n-2}u_2 + \cdots + b_1u_{n-1} + b_0u_n. \tag{5.1}$$

We are thus in possession of an explicit expression for v_n. [For an alternative proof see example (10.*a*).] The relations (5.1) may be rewritten in the compact form of a convolution equation:

$$\{v_n\} = \{b_n\} * \{u_n\}. \tag{5.2}$$

This implies that the corresponding generating functions satisfy the identity

$$V(s) = B(s)U(s) = \frac{B(s)}{1 - F(s)}. \tag{5.3}$$

Example. (*a*) In the Bernoulli trials considered in examples (4.*a*)–(4.*d*) the event $\mathbf{S}_n = 1$ is a delayed recurrent event. The waiting time for its first occurrence has the generating function Φ of (4.10); the recurrence times between successive occurrences of $\{\mathbf{S}_n = 1\}$ have the generating function F of the returns to equilibrium [see (4.4)]. Thus in the present case $V = \Phi/(1-F)$. ▶

It is easy to show that the asymptotic behavior of the probabilities v_n is essentially the same as that of u_n. To avoid trivialities we assume that $\mathcal{E}$ is not periodic.[5] We know from section 3 that in this case u_n approaches a finite limit, and that $\sum u_n < \infty$ if, and only if, $\mathcal{E}$ is transient.

Theorem 1. *If* $u_n \to \omega$ *then*

$$v_n \to b\omega \qquad \textit{where} \quad b = \sum b_k = B(1). \tag{5.4}$$

If $\sum u_n = u < \infty$ *then*

$$\sum v_n = bu. \tag{5.5}$$

In particular, $v_n \to \mu^{-1}$ *if* $\mathcal{E}$ *is persistent.*

Proof. Let $r_k = b_{k+1} + b_{k+2} + \cdots$. Since $u_n \leq 1$ it is obvious from (5.1) that for $n > k$

$$b_0u_n + \cdots + b_ku_{n-k} \leq v_n \leq b_0u_n + \cdots + b_ku_{n-k} + r_k. \tag{5.6}$$

Choose k so large that $r_k < \epsilon$. For n sufficiently large the leftmost member in (5.6) is then greater than $b\omega - 2\epsilon$, whereas the rightmost member is less than $b\omega + 2\epsilon$. Thus (5.4) is true. The assertion (5.5) follows either by summing (5.1) over n, or else from (5.3) on letting $s = 1$. ►

We turn to a general limit theorem of wide applicability. Suppose that there are denumerably many possible states $E_0, E_1, \ldots$ for a certain system, and that the transitions from one state to another depend on a chance mechanism of some sort. For example, in the simple queuing process (1.*h*) we say that the system is in state E_k if there are k customers in the queue, including the customer being served. A problem involving seventeen servers may require eighteen numbers to specify the state of the system, but all imaginable states can still be ordered in a sequence $E_0, E_1, \ldots$. We need not consider how this is best done, because the following theorem does not lead to practical methods for evaluating probabilities. It is a pure existence theorem showing that a steady state exists under most circumstances encountered in practice. This is of conceptual interest, but also of practical value because, as a rule, mathematical analysis of a steady state is much simpler than the study of the time-dependent process.

We suppose that for $n = 1, 2, \ldots$ and every n-tuple $(r_1, \ldots, r_n)$ there exists a well-defined probability that the states of the system at epochs

[5] Periodic recurrent events are covered by theorem 10.2. For a different proof of theorem 1 see example (10.*a*).

$0, 1, \ldots, n - 1$ are represented by $(E_{r_1}, \ldots, E_{r_n})$. We shall not introduce any particular assumptions concerning the mutual dependence of these events or the probabilities for the transitions from one state to another. For simplicity we consider only *the probabilities* $p_n^{(r)}$ *that at epoch* n *the system is in state* E_r. (It will be obvious how the theorem generalizes to pairs, triples, etc.) The crucial assumption is that there exists some recurrent event $\mathcal{E}$ connected with our process. For example, in the queuing process (1.*h*) the state E_0 represents such a recurrent event. In this case, if $\mathcal{E}$ were transient there would exist a positive probability that the queue does not terminate. This would imply that sooner or later we would encounter an unending queue, that is, a queue of indefinitely increasing size. This is a limit theorem of some sort showing that such servers are impossible in practice. This example should explain the role of the condition that $\mathcal{E}$ be persistent. (The non-periodicity is introduced only to avoid trivialities).

Theorem 2. *Assume that there exists a non-periodic persistent (possibly delayed) recurrent event* $\mathcal{E}$ *associated with our process. Then as* $n \to \infty$

$$p_n^{(r)} \to p^{(r)} \tag{5.7}$$

where

$$\sum p^{(r)} = 1 \tag{5.8}$$

if the mean recurrence time μ *is finite, and* $p^{(r)} = 0$ *otherwise.*

Proof. Every time when $\mathcal{E}$ occurs the process starts from scratch. There exists therefore a well-defined conditional probability $g_n^{(r)}$ that if $\mathcal{E}$ occurs at some epoch, the state E_r occurs n time units later and *before* the next occurrence of $\mathcal{E}$ (here $n = 0, 1, \ldots$). For delayed recurrent events we require also the probability $\gamma_n^{(r)}$ that E_r occurs at epoch n *before* the first occurrence of $\mathcal{E}$. (Clearly $\gamma_n^{(r)} = g_n^{(r)}$ if $\mathcal{E}$ is not delayed.)

Let us now classify the ways in which E_r can occur at epoch n according to the last occurrence of $\mathcal{E}$ before epoch n. First, it is possible that $\mathcal{E}$ did not yet occur. The probability for this is $\gamma_n^{(r)}$. Or else there exists a $k \leq n$ such that $\mathcal{E}$ occurred at epoch k but not between k and n. The probability for this is $v_kg_{n-k}^{(r)}$. Summing over all mutually exclusive cases we find

$$p_n^{(r)} = \gamma_n^{(r)} + g_{n-1}^{(r)}v_1 + g_{n-2}^{(r)}v_2 + \cdots + g_0^{(r)}v_n. \tag{5.9}$$

(Here we adhere to the notations of theorem 1. For delayed events $v_0 = 0$; for non-delayed events $v_k = u_k$ and $\gamma_n^{(r)} = g_n^{(r)}$.)

The relation (5.9) is analogous to (5.1) except for the appearance of the term $\gamma_n^{(r)}$ on the right. This quantity is obviously smaller than the probability that $\mathcal{E}$ did not occur before epoch n, and $\mathcal{E}$ being persistent it follows that $\gamma_n^{(r)} \to 0$ as $n \to \infty$. For the remaining terms we can apply theorem 1 with the notational change that u_k is replaced by v_k and b_k by $g_n^{(r)}$. Since $\mathcal{E}$ is persistent $v_n \to \mu^{-1}$ and it follows that

$$p_n^{(r)} \to \mu^{-1} \sum_{k=0}^{\infty} g_k^{(r)}. \tag{5.10}$$

This proves the existence of the limits (5.7). To prove that they add to unity note that at any epoch the system is in some state and hence

$$\sum_{r=0}^{\infty} g_n^{(r)} = g_n \tag{5.11}$$

is the probability that a recurrence time is $\geq n$, that is,

$$g_n = f_n + f_{n+1} + \cdots.$$

Thus

$$\sum_{r=0}^{\infty} p^{(r)} = \frac{1}{\mu} \sum_{n=0}^{\infty} g_n = 1 \tag{5.12}$$

by XI, (1.8). ▶

[The limit theorem in example (10.*b*) may be treated as a special case of the present theorem.]

6. THE NUMBER OF OCCURRENCES OF $\mathcal{E}$

Up to now we have studied a recurrent event $\mathcal{E}$ in terms of the waiting times between its successive occurrences. Often it is preferable to consider the number n of trials as given and to take *the number* $\mathbf{N}_n$ *of occurrences of* $\mathcal{E}$ in the first n trials as basic variable. We shall now investigate the asymptotic behavior of the distribution of $\mathbf{N}_n$ for large n. For simplicity we assume that $\mathcal{E}$ is not delayed.

As in (2.8) let $\mathbf{T}^{(r)}$ stand for the number of trials up to and including the rth occurrence of $\mathcal{E}$. The probability distributions of $\mathbf{T}^{(r)}$ and $\mathbf{N}_n$ are related by the obvious identity

$$\mathbf{P}\{\mathbf{N}_n \geq r\} = \mathbf{P}\{\mathbf{T}^{(r)} \leq n\}. \tag{6.1}$$

We begin with the simple case where $\mathcal{E}$ is persistent and the distribution $\{f_n\}$ of its recurrence times has finite mean μ and variance σ^2. Since $\mathbf{T}^{(r)}$ is the sum of r independent variables, the central limit theorem of X,1 asserts that for each fixed x as $r \to \infty$

$$\mathbf{P}\left\{\frac{\mathbf{T}^{(r)} - r\mu}{\sigma\sqrt{r}} < x\right\} \to \mathfrak{N}(x) \tag{6.2}$$

where $\mathfrak{N}(x)$ is the normal distribution function. Now let $n \to \infty$ and $r \to \infty$ in such a way that

$$\frac{n - r\mu}{\sigma\sqrt{r}} \to x; \tag{6.3}$$

then (6.1) and (6.2) together lead to

$$\mathbf{P}\{\mathbf{N}_n \geq r\} \to \mathfrak{N}(x). \tag{6.4}$$

To write this relation in a more familiar form we introduce the *reduced variable*

$$\mathbf{N}_n^* = (\mu \mathbf{N}_n - n)\sqrt{\frac{\mu}{\sigma^2 n}}. \tag{6.5}$$

The inequality $\mathbf{N}_n \geq r$ is identical with

$$\mathbf{N}_n^* \geq \frac{r\mu - n}{\sigma\sqrt{r}} \cdot \sqrt{\frac{r\mu}{n}} = -x\sqrt{\frac{r\mu}{n}}. \tag{6.6}$$

On dividing (6.3) by r it is seen that $n/r \to \mu$, and hence the right side in (6.6) tends to $-x$. Since $\mathfrak{N}(-x) = 1 - \mathfrak{N}(x)$ it follows that

$$\mathbf{P}\{\mathbf{N}_n^* \geq -x\} \to \mathfrak{N}(x) \quad \text{or} \quad \mathbf{P}\{\mathbf{N}_n^* < -x\} \to 1 - \mathfrak{N}(x), \tag{6.7}$$

and we have proved the

Theorem. *Normal approximation. If the recurrent event* $\mathcal{E}$ *is persistent and its recurrence times have finite mean* μ *and variance* σ^2, *then both the number* $\mathbf{T}^{(r)}$ *of trials up to the rth occurrence of* $\mathcal{E}$ *and the number* $\mathbf{N}_n$ *of occurrences of* $\mathcal{E}$ *in the first* n *trials are asymptotically normally distributed as indicated in* (6.2) *and* (6.7).

Note that in (6.7) we have the central limit theorem applied to a sequence of *dependent* variables $\mathbf{N}_n$. Its usefulness will be illustrated in the next section by an application to the theory of runs.

The relations (6.7) make it plausible that

$$\mathbf{E}(\mathbf{N}_n) \sim n/\mu, \qquad \mathrm{Var}\,(\mathbf{N}_n) \sim n\sigma^2/\mu^3 \tag{6.8}$$

where the sign $\sim$ indicates that the ratio of the two sides tends to unity. To prove (6.8) we note that $\mathbf{N}_n$ is the sum of n (dependent) variables $\mathbf{Y}_k$

such that $\mathbf{Y}_k$ equals one or zero according as $\mathcal{E}$ does or does not occur at the kth trial. Thus $\mathbf{E}(\mathbf{Y}_k) = u_k$ and

$$\mathbf{E}(\mathbf{N}_n) = u_1 + u_2 + \cdots + u_n. \tag{6.9}$$

Since $u_n \to \mu^{-1}$ this implies the first relation in (6.8). The second follows by a similar argument (see problem 20).

Unfortunately surprisingly many recurrence times occurring in various stochastic processes and in applications have *infinite expectations.* In such cases the normal approximation is replaced by more general limit theorems of an entirely different character,[6] and the chance fluctuations exhibit unexpected features. For example, one expects intuitively that $\mathbf{E}(\mathbf{N}_n)$ should increase linearly with n "because on the average $\mathcal{E}$ must occur twice as often in twice as many trials." Yet *this is not so.* An infinite mean recurrence time implies that $u_n \to 0$, and then $\mathbf{E}(\mathbf{N}_n)/n \to 0$ by virtue of (6.9). This means that in the long run the occurrences of $\mathcal{E}$ become rarer and rarer, and this is possible only if some recurrence times are fantastically large. Two examples may show how pronounced this phenomenon is apt to be.

Examples. (*a*) When $\mathcal{E}$ stands for a return to equilibrium in a coin-tossing game [example (4.*b*) with $p = \frac{1}{2}$] we have $u_{2n} \sim 1/\sqrt{\pi n}$, and (6.9) approximates an integral for $(\pi x)^{-\frac{1}{2}}$; this implies $\mathbf{E}(\mathbf{N}_{2n}) \sim 2\sqrt{n/\pi}$. Thus the *average* recurrence time up to epoch n is likely to increase as $\sqrt{n}$. The curious consequences of this were discussed at length in chapter III.

(*b*) Returning to example (4.*g*) consider repeated tosses of $r = 3$ dice and let $\mathcal{E}$ stand for the event that all three coins are in the same phase. We saw that $\mathcal{E}$ is a persistent recurrent event, and that $u_n \sim \dfrac{2}{\sqrt{3} \cdot \pi n}$. Thus $\mathbf{E}(\mathbf{N}_n)$ *increases roughly* as $\log n$ and so the *average* of the recurrence times up to epoch n is likely to be of the fantastic magnitude $n/\log n$. ▶

*7. APPLICATION TO THE THEORY OF SUCCESS RUNS

In the sequel r will denote a fixed positive integer and $\mathcal{E}$ will stand for the occurrence of a success run of length r in a sequence of Bernoulli trials. It is important that the length of a run be defined as stated in

* Sections 7 and 8 treat a special topic and may be omitted.

[6] W. Feller, *Fluctuation theory of recurrent events*, Trans. Amer. Math. Soc., vol. 67 (1949), pp. 98–119.

example (1.*e*), for otherwise runs are not recurrent events, and the calculations become more involved. As in (2.1) and (2.2), u_n *is the probability of* $\mathcal{E}$ *at the nth trial, and* f_n *is the probability that the first run of length* r *occurs at the nth trial.*

The probability that the r trials number $n, n-1, n-2, \ldots, n-r+1$ result in success is obviously p^r. In this case $\mathcal{E}$ occurs at one among these r trials; the probability that $\mathcal{E}$ occurs at the trial number $n-k$ $(k = 0, 1, \ldots, r-1)$ and the following k trials result in k successes equals $u_{n-k}p^k$. Since these r possibilities are mutually exclusive, we get the recurrence relation[7]

$$u_n + u_{n-1}p + \cdots + u_{n-r+1}p^{r-1} = p^r \tag{7.1}$$

valid for $n \geq r$. Clearly

$$u_1 = u_2 = \cdots = u_{r-1} = 0, \qquad u_0 = 1. \tag{7.2}$$

On multiplying (7.1) by s^n and summing over $n = r,\ r+1,\ r+2, \ldots,$ we get on the left side

$$\{U(s) - 1\}(1+ps+p^2s^2+ \cdots +p^{r-1}s^{r-1}) \tag{7.3}$$

and on the right side $p^r(s^r+s^{r+1}+ \cdots)$. The two series are geometric, and we find that

$$\{U(s) - 1\} \cdot \frac{1-(ps)^r}{1-ps} = \frac{p^rs^r}{1-s} \tag{7.4}$$

or

$$U(s) = \frac{1-s+qp^rs^{r+1}}{(1-s)(1-p^rs^r)}. \tag{7.5}$$

From (3.2), we get now *the generating function of the recurrence times:*

$$F(s) = \frac{p^rs^r(1-ps)}{1-s+qp^rs^{r+1}} = \frac{p^rs^r}{1-qs(1+ps+ \cdots +p^{r-1}s^{r-1})}. \tag{7.6}$$

The fact that $F(1) = 1$ shows that in a prolonged sequence of trials the number of runs of any length is certain to increase over all bounds. The mean recurrence time μ could be obtained directly from (7.1) since we know that $u_n \to \mu^{-1}$. Since we require also the variance, it is preferable

[7] The classical approach consists in deriving a recurrence relation for f_n. This method is more complicated and does not apply to, say, runs of either kind or patterns like *SSFFSS*, to which our method applies without change [cf. example (8.*c*)].

to calculate the derivatives of $F(s)$. This is best done by implicit differentiation after clearing (7.6) of the denominator. An easy calculation then shows that *the mean and variance of the recurrence times of runs of length r are*

$$\mu = \frac{1-p^r}{qp^r}, \qquad \sigma^2 = \frac{1}{(qp^r)^2} - \frac{2r+1}{qp^r} - \frac{p}{q^2}, \tag{7.7}$$

respectively. By the theorem of the last section for large n *the number* N_n *of runs of length r produced in n trials is approximately normally*

TABLE 2

MEAN RECURRENCE TIMES FOR SUCCESS RUNS IF TRIALS ARE PERFORMED AT THE RATE OF ONE PER SECOND

Length of Run	$p = 0.6$	$p = 0.5$ (Coins)	$p = \frac{1}{6}$ (Dice)
$r = 5$	30.7 seconds	1 minute	2.6 hours
10	6.9 minutes	34.1 minutes	28.0 months
15	1.5 hours	18.2 hours	18,098 years
20	19 hours	24.3 days	140.7 million years

distributed, that is, for fixed $\alpha < \beta$ the probability that

$$\frac{n}{\mu} + \alpha\sigma\sqrt{\frac{n}{\mu^3}} < \mathrm{N}_n < \frac{n}{\mu} + \beta\sigma\sqrt{\frac{n}{\mu^3}} \tag{7.8}$$

tends to $\mathfrak{N}(\beta) - \mathfrak{N}(\alpha)$. This fact was first proved by von Mises, by rather lengthy calculations. Table 2 gives a few typical means of recurrence times.

The method of partial fractions of XI, 4, permits us to derive excellent approximations. The second representation in (7.6) shows clearly that the denominator has a unique *positive root* $s = x$. For every real or complex number s with $|s| \leq x$ we have

$$|qs(1+ps+\cdots+p^{r-1}s^{r-1})| \leq qx(1+px+\cdots+p^{r-1}x^{r-1}) = 1 \tag{7.9}$$

where the equality sign is possible only if all terms on the left have the same argument, that is, if $s = x$. Hence x is smaller in absolute value than any other root of the denominator in (7.6). We can, therefore, apply formulas (4.5) and (4.9) of chapter XI with $s_1 = x$, letting $U(s) = p^rs^r(1-ps)$ and $V(s) = 1 - s + qp^rs^{r+1}$. We find, using that $V(x) = 0$,

$$f_n \sim \frac{(x-1)(1-px)}{(r+1-rx)q} \cdot \frac{1}{x^{n+1}}. \tag{7.10}$$

The probability of no run in n trials is $q_n = f_{n+1} + f_{n+2} + f_{n+3} + \cdots$ and summing the geometric series in (7.10) we get

$$q_n \sim \frac{1-px}{(r+1-rx)q} \cdot \frac{1}{x^{n+1}}. \tag{7.11}$$

We have thus found that *the probability of no success run of length r in n trials satisfies* (7.11). Table 3 shows that the right side gives surprisingly good approximations even for very small n, and the approximation improves rapidly with n. This illustrates the power of the method of generating function and partial fractions.

TABLE 3

PROBABILITY OF HAVING NO SUCCESS RUN OF LENGTH $r = 2$ IN n TRIALS WITH $p = \frac{1}{2}$

n	q_n Exact	From (7.11)	Error
2	0.75	0.76631	0.0163
3	0.625	0.61996	0.0080
4	0.500	0.50156	0.0016
5	0.40625	0.40577	0.0005

Numerical Calculations. For the benefit of the practical-minded reader we use this occasion to show that the numerical calculations involved in partial fraction expansions are often less formidable than they appear at first sight, and that excellent estimates of the error can be obtained.

The asymptotic expansion (7.11) raises two questions: First, the contribution of the $r - 1$ neglected roots must be estimated, and second, the dominant root x must be evaluated.

The first representation in (7.6) shows that all roots of the denominator of $F(s)$ satisfy the equation

$$s = 1 + qp^rs^{r+1}, \tag{7.12}$$

but (7.12) has the additional extraneous root $s = p^{-1}$. For positive s the graph of $f(s) = 1 + qp^rs^{r+1}$ is convex; it intersects the bisector $y = s$ at x and p^{-1} and in the interval between x and p^{-1} the graph lies *below* the bisector. Furthermore, $f'(p^{-1}) = (r+1)q$. If this quantity exceeds unity, the graph of $f(s)$ crosses the bisector at $s = p$ from below, and hence $p^{-1} > x$. To fix ideas we shall assume that

$$(r+1)q > 1; \tag{7.13}$$

in this case $x < p^{-1}$, and $f(s) < s$ for $x < s < p^{-1}$. It follows that for all complex numbers s such that $x < |s| < p^{-1}$ we have $|f(s)| \leq f(|s|) < |s|$ so that no root s_k can lie in the annulus $x < |s| < p^{-1}$. Since x was chosen as the root smallest in absolute value, this implies that

$$(7.14) \qquad |s_k| > p^{-1} \qquad \textit{when} \quad s_k \neq x.$$

By differentiation of (7.12) it is now seen that all roots are simple.

The contribution of each root to q_n is of the same form as the contribution (7.11) of the dominant root x, and therefore the $r - 1$ terms neglected in (7.11) are of the form

$$(7.15) \qquad A_k = \frac{ps_k - 1}{rs_k - (r+1)} \cdot \frac{1}{qs_k^{n+1}}.$$

We require an upper bound for the first fraction on the right. For that purpose note that for fixed $s > p^{-1} > (r+1)r^{-1}$

$$(7.16) \qquad \left|\frac{pse^{i\theta} - 1}{rse^{i\theta} - (r+1)}\right| \leq \frac{ps+1}{rs+r+1};$$

in fact, the quantity on the left obviously assumes its maximum and minimum for $\theta = 0$ and $\theta = \pi$, and a direct substitution shows that 0 corresponds to a minimum, π to a maximum. In view of (7.13) and (7.14) we have then

$$(7.17) \qquad |A_k| < \frac{2p^{n+1}}{(r+1+rp^{-1})q} < \frac{2p^{n+2}}{rq(1+p)}.$$

We conclude that in (7.11) *the error committed by neglecting the* $r - 1$ *roots different from* x *is less in absolute value than*

$$(7.18) \qquad \frac{2(r-1)p}{rq(1+p)}.$$

The root x is easily calculated from (7.12) by successive approximations putting $x_0 = 1$ and $x_{\nu+1} = f(x_\nu)$. The sequence will converge monotonically to x, and each term provides a lower bound for x, whereas any value s such that $s > f(s)$ provides an upper bound. It is easily seen that

$$(7.19) \qquad x = 1 + qp^r + (r+1)(qp^r)^2 + \cdots.$$

*8. MORE GENERAL PATTERNS

Our method is applicable to more general problems which have been considered as considerably more difficult than simple runs.

Examples. (*a*) *Runs of either kind.* Let $\mathcal{E}$ stand for "*either a success run of length* r *or a failure run of length* ρ" [see example (1.*f*)]. We are dealing with *two* recurrent events $\mathcal{E}_1$ and $\mathcal{E}_2$, where $\mathcal{E}_1$ stands for "success run of length r" and $\mathcal{E}_2$ for "failure run of length ρ" and $\mathcal{E}$ means "either $\mathcal{E}_1$ or $\mathcal{E}_2$." To $\mathcal{E}_1$ there corresponds the generating function (7.5) which will now be denoted by $U_1(s)$. The corresponding generating function

$U_2(s)$ for $\mathcal{E}_2$ is obtained from (7.5) by interchanging p and q and replacing r by ρ. The probability u_n that $\mathcal{E}$ occurs at the nth trial is the sum of the corresponding probabilities for $\mathcal{E}_1$ and $\mathcal{E}_2$, except that $u_0 = 1$. It follows that

$$(8.1) \qquad U(s) = U_1(s) + U_2(s) - 1.$$

The generating function $F(s)$ of the recurrence times of $\mathcal{E}$ is again $F(s) = 1 - U^{-1}(s)$ or

$$(8.2) \qquad F(s) = \frac{(1-ps)p^r s^r(1-q^\rho s^\rho) + (1-qs)q^\rho s^\rho(1-p^r s^r)}{1 - s + qp^r s^{r+1} + pq^\rho s^{\rho+1} - p^r q^\rho s^{r+\rho}}.$$

The *mean recurrence time* follows by differentiation

$$(8.3) \qquad \mu = \frac{(1-p^r)(1-q^\rho)}{qp^r + pq^\rho - p^r q^\rho}.$$

As $\rho \to \infty$, this expression tends to the mean recurrence time of success runs as given in (7.7).

(*b*) In VIII,1, we calculated the probability x that a *success run of length* r *occurs before a failure run of length* ρ. Define two recurrent events $\mathcal{E}_1$ and $\mathcal{E}_2$ as in example (*a*). Let x_n = probability that $\mathcal{E}_1$ occurs for the first time at the nth trial and no $\mathcal{E}_2$ precedes it; f_n = probability that $\mathcal{E}_1$ occurs for the first time at the nth trial (with no condition on $\mathcal{E}_2$). Define y_n and g_n as x_n and f_n, respectively, but with $\mathcal{E}_1$ and $\mathcal{E}_2$ interchanged.

The generating function for f_n is given in (7.6), and $G(s)$ is obtained by interchanging p and q and replacing r by ρ. For x_n and y_n we have the obvious recurrence relations

$$(8.4) \qquad \begin{aligned} x_n &= f_n - (y_1 f_{n-1} + y_2 f_{n-2} + \cdots + y_{n-1} f_1) \\ y_n &= g_n - (x_1 g_{n-1} + x_2 g_{n-2} + \cdots + x_{n-1} g_1). \end{aligned}$$

They are of the convolution type, and for the corresponding generating functions we have, therefore,

$$(8.5) \qquad \begin{aligned} X(s) &= F(s) - Y(s)F(s) \\ Y(s) &= G(s) - X(s)G(s). \end{aligned}$$

From these two linear equations we get

$$(8.6) \qquad X(s) = \frac{F(s)\{1 - G(s)\}}{1 - F(s)G(s)}, \qquad Y(s) = \frac{G(s)\{1 - F(s)\}}{1 - F(s)G(s)}.$$

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

Expressions for x_n and y_n can again be obtained by the method of partial fractions. For $s = 1$ we get $X(1) = \sum x_n = x$, the probability of $\mathcal{E}_1$ occurring before $\mathcal{E}_2$. Both numerator and denominator vanish, and $X(1)$ is obtained from L'Hospital's rule differentiating numerator and denominator: $X(1) = G'(1)/\{F'(1) + G'(1)\}$. Using the values $F'(1) = (1-p^r)/qp^r$ and $G'(1) = (1-q^\rho)/pq^\rho$ from (7.7), we find $X(1)$ as given in VIII,(1.3).

(*c*) Consider the recurrent event defined by the pattern *SSFFSS*. Repeating the argument of section 7, we easily find that

$$p^4q^2 = u_n + p^2q^2u_{n-4} + p^3q^2u_{n-5}. \tag{8.7}$$

Since we know that $u_n \to \mu^{-1}$ we get for the mean recurrence time $\mu = p^{-4}q^{-2} + p^{-2} + p^{-1}$. For $p = q = \frac{1}{2}$ we find $\mu = 70$, whereas the mean recurrence time for a success run of length 6 is 126. This shows that, contrary to expectation, *there is an essential difference in coin tossing between head runs and other patterns of the same length.* ▶

9. LACK OF MEMORY OF GEOMETRIC WAITING TIMES

The geometric distribution for waiting times has an interesting and important property not shared by any other distribution. Consider a sequence of Bernoulli trials and let $\mathbf{T}$ be the number of trials up to and including the first success. Then $\mathbf{P}\{\mathbf{T} > k\} = q^k$. Suppose we know that no success has occurred during the first m trials; the waiting time $\mathbf{T}$ from this mth failure to the first success has exactly the same distribution $\{q^k\}$ and is independent of the number of preceding failures. In other words, the probability that the waiting time will be prolonged by k always equals the initial probability of the total length exceeding k. If the life span of an atom or a piece of equipment has a geometric distribution, then *no aging* takes place; as long as it lives, the atom has the same probability of decaying at the next trial. Radioactive atoms actually have this property (except that in the case of a continuous time the exponential distribution plays the role of the geometric distribution). Conversely, if it is known that a phenomenon is characterized by a complete lack of memory or aging, then the probability distribution of the duration must be geometric or exponential. Typical is a well-known type of telephone conversation often cited as the model of incoherence and depending entirely on momentary impulses; a possible termination is an instantaneous chance effect without relation to the past chatter. By contrast, the knowledge that no streetcar has passed for five minutes increases our expectation that it will come soon. In coin tossing, the probability that the cumulative numbers of heads and tails will equalize at the second trial is $\frac{1}{2}$. However, given that they did not, the probability that they equalize after two additional trials is only $\frac{1}{4}$. These are examples for aftereffect.

For a rigorous formulation of the assertion, suppose that a waiting time $\mathbf{T}$ assumes the values $0, 1, 2, \ldots$ with probabilities $p_0, p_1, p_2, \ldots$. Let the distribution of $\mathbf{T}$ have the following property: *The conditional probability that the waiting time terminates at the kth trial, assuming that it has not terminated before, equals p_0 (the probability at the first trial). We claim that $p_k = (1-p_0)^kp_0$, so that $\mathbf{T}$ has a geometric distribution.*

For a proof we introduce again the "tails"

$$q_k = p_{k+1} + p_{k+2} + p_{k+3} + \cdots = \mathbf{P}\{\mathbf{T} > k\}.$$

Our hypothesis is $\mathbf{T} > k - 1$, and its probability is q_{k-1}. The conditional probability of $\mathbf{T} = k$ is therefore p_k/q_{k-1}, and the assumption is that for all $k \geq 1$

$$\frac{p_k}{q_{k-1}} = p_0. \tag{9.1}$$

Now $p_k = q_{k-1} - q_k$, and hence (9.1) reduces to

$$\frac{q_k}{q_{k-1}} = 1 - p_0. \tag{9.2}$$

Since $q_0 = p_1 + p_2 + \cdots = 1 - p_0$, it follows that $q_k = (1-p_0)^{k+1}$, and hence $p_k = q_{k-1} - q_k = (1 - p_0)^kp_0$, as asserted. ▶

In the theory of stochastic processes the described lack of memory is connected with the *Markovian property;* we shall return to it in XV,13.

10. RENEWAL THEORY

The convolution equations which served as a basis for the theory of recurrent events are of much wider applicability than appears in the foregoing sections. We shall therefore restate their analytic content in somewhat greater generality and describe the typical probabilistic renewal argument as well as applications to the study of populations of various sorts.

We start from two arbitrary sequences[8] $f_1, f_2, \ldots$ and $b_0, b_1, \ldots$ of real numbers. A new sequence $v_0, v_1, \ldots$ may then be defined by the

[8] We put $f_0 = 0$. It is clear from (10.1) that the case $0 < f_0 < 1$ involves only the change of notations, replacing f_k by $f_k \,|\, (1 - f_0)$ and b_k by $b_k \,|\, (1 - f_0)$.

convolution equations

$$v_n = b_n + f_1 v_{n-1} + f_2 v_{n-2} + \cdots + f_n v_0. \tag{10.1}$$

These define recursively $v_0, v_1, v_2, \ldots$ and so the v_n are uniquely defined under any circumstances. We shall, however, consider only sequences satisfying the conditions[9]

$$f_n \geq 0, \qquad f = \sum_{n=1}^{\infty} f_n < \infty; \qquad b_n \geq 0, \qquad b = \sum_{n=0}^{\infty} b_n < \infty. \tag{10.2}$$

In this case the v_n are non-negative and the corresponding generating functions must satisfy the identity

$$V(s) = \frac{B(s)}{1 - F(s)}. \tag{10.3}$$

The generating functions F and B converge at least for $0 \leq s < 1$, and so (10.3) defines a power series converging as long as $F(s) < 1$. Relations (10.1) and (10.3) are fully equivalent. In section 3 we considered the special case $B(s) = 1$ (with $v_n = u_n$ for all n). Section 5 covered the general situation under the restriction $f \leq 1$. In view of applications to population theory we shall now permit that $f > 1$; fortunately this case is easily reduced to the standard case $f = 1$.

We shall say that *the sequence* $\{f_n\}$ *has period* $\lambda > 1$ if $f_n = 0$ except when $n = k\lambda$ is a multiple of λ, and λ is the greatest integer with this property. This amounts to saying that $F(s) = F_1(s^\lambda)$ is a power series in s^λ, but not in $s^{r\lambda}$ for any $r > 1$. We put again

$$\mu = \sum n f_n \leq \infty \tag{10.4}$$

and adhere to the convention that μ^{-1} is to be interpreted as 0 if $\mu = \infty$.

Theorem 1. (*Renewal theorem.*) *Suppose* (10.2) *and that* $\{f_n\}$ *is not periodic.*

(i) *If* $f < 1$ *then* $v_n \to 0$ *and*

$$\sum_{n=0}^{\infty} v_n = \frac{b}{1 - f}. \tag{10.5}$$

(ii) *If* $f = 1$

$$v_n \to b\mu^{-1}. \tag{10.6}$$

[9] The positivity of f_n is essential, but the convergence of the two series is imposed only for convenience. No general conclusion can be drawn if $b = \infty$ *and* $f = \infty$. The assertion (10.7) remains true when $f = \infty$ except that in this case $F'(\xi)$ is not necessarily finite, and (10.7) is meaningless if $b = \infty$ *and* $F'(\xi) = \infty$.

(iii) *If* $f > 1$ *there exists a unique positive root of the equation* $F(\xi) = 1$, *and*

$$\xi^n v_n \to \frac{B(\xi)}{\xi F'(\xi)}. \tag{10.7}$$

Obviously $\xi < 1$ and hence the derivative $F'(\xi)$ is finite; (10.7) shows that the sequence $\{v_n\}$ behaves ultimately like a geometric sequence with ratio $\xi^{-1} > 1$.

Proof. The assertions (i) and (ii) were proved in section 5. To prove (iii) it suffices to apply the result (ii) to the sequences $\{f_n\xi^n\}$, $\{b_n\xi^n\}$, and $\{v_n\xi^n\}$ with generating functions given by $F(\xi s)$, $B(\xi s)$, and $V(\xi s)$, respectively. ▶

We have excluded periodic sequences $\{f_n\}$ because they are of secondary interest. Actually they present nothing new. Indeed, if $\{b_n\}$ and $\{f_n\}$ have the same period λ then both $B(s)$ and $F(s)$ are power series in s^λ, and hence the same is true of $V(s)$. Theorem 1 then applies to the sequences $\{f_{n\lambda}\}$, $\{b_{n\lambda}\}$, and $\{v_{n\lambda}\}$ with generating functions $F(s^{1/\lambda})$, $B(s^{1/\lambda})$, and $V(s^{1/\lambda})$. When $F(1) = 1$ it follows that $v_{n\lambda} \to b\lambda/\mu$. Now the most general power series B can be written as a linear combination

$$B(s) = B_0(s) + sB_1(s) + \cdots + s^{\lambda-1}B_{\lambda-1}(s) \tag{10.8}$$

of λ power series B_j each of which involves only powers of s^λ. Introducing this into (10.3) and applying the result just stated shows the validity of

Theorem 2. *Let* (10.2) *hold and suppose that* $\{f_n\}$ *has period* $\lambda > 1$.

(i) *If* $f < 1$ *then* (10.5) *holds.*
(ii) *If* $f = 1$ *then for* $j = 0, 1, \ldots, \lambda - 1$ *as* $n \to \infty$

$$u_{n\lambda+j} \to \lambda B_j(1)/\mu. \tag{10.9}$$

(iii) *If* $f > 1$ *then for* $j = 0, 1, \ldots, \lambda - 1$ *as* $n \to \infty$

$$\xi^{n\lambda} u_{n\lambda+j} \to \lambda B_j(\xi)/(\xi\mu). \tag{10.10}$$

In a great variety of stochastic processes it is possible to adapt the argument used for recurrent events to show that certain probabilities satisfy an equation of the convolution type like (10.1). Many important limit theorems appear in this way as simple corollaries of theorem 1. This approach has now generally supplanted clumsier older methods and has become known as *renewal argument*. Its full power appears only when used for processes with a continuous time parameter, but the first two examples may serve as an illustration. For further examples see problems 8–9. An application of theorem 1 to a non-probabilistic limit theorem is contained in example (*c*). The last two examples are devoted to practical applications.

Examples. (*a*) *Delayed recurrent events.* We give a new derivation of the result in section 5 for a delayed recurrent event $\mathcal{E}$ with the distribution $\{f_j\}$ for the recurrence times, and the distribution $\{b_j\}$ for the *first* occurrence of $\mathcal{E}$. Let v_n stand for the probability that $\mathcal{E}$ occurs at the nth trial. We show that (10.1) holds. There are two ways in which $\mathcal{E}$ can occur at the nth trial. The occurrence may be the first, and the probability for this is b_n. Otherwise there was a last occurrence of $\mathcal{E}$ before the nth trial, and so there exists a number $1 \leq j < n$ such that $\mathcal{E}$ did occur at the jth trial and the *next* time at the nth trial. The probability for this is $v_j f_{n-j}$. The cases are mutually exclusive, and so

$$v_n = b_n + v_1 f_{n-1} + v_2 f_{n-2} + \cdots + v_{n-1} f_1, \tag{10.11}$$

which is the same as (10.1). The generating function V is therefore given by (10.3) in agreement with the result in section 5. (Though the results agree even formally, the arguments are different: in section 5 the enumeration proceeded according to the first appearance of $\mathcal{E}$ whereas the present argument uses the last appearance. Both procedures are used in other circumstances and sometimes lead to formally different equations.)

(*b*) *Hitting probabilities.* Consider a sequence of trials with a proper (not delayed) persistent recurrent event $\mathcal{E}$. Let $\nu \geq 0$ be an integer. Suppose that we start to observe the process only after the νth trial and that we are interested in the waiting time for the next occurrence of $\mathcal{E}$. More formally, for $r = 1, 2, \ldots$ denote by $w_\nu(r)$ the probability that *the first occurrence of* $\mathcal{E}$ *after the* νth *trial* takes place at the $(\nu+r)$th trial. Thus $w_0(r) = f_r$ and $w_\nu(0) = 0$. [The $w_\nu(r)$ are called hitting probabilities because of their meaning in random walks. In other contexts it is more natural to speak of the distribution of the residual waiting time commencing at the νth trial. Cf. example XV,(2.k).]

To determine these probabilities we use the standard renewal argument as follows. It is possible that $\mathcal{E}$ occurs for the very first time at the $(\nu+r)$th trial. The probability for this is $f_{\nu+r}$. Otherwise there exists an integer $k \leq \nu$ such that $\mathcal{E}$ occurred for the first time at the kth trial. The continuation of the process after the kth trial is a probabilistic replica of the whole process, except that the original νth trial now becomes the $(\nu-k)$th trial. The probability of our event is therefore $f_k w_{\nu-k}(r)$, and hence for each $r > 0$

$$w_\nu(r) = f_{\nu+r} + \sum_{k=1}^{\nu} f_k w_{\nu-k}(r). \tag{10.12}$$

This equation is of the standard type (10.1) with $b_n = f_{n+r}$. We are not interested in the generating function but wish to describe the asymptotic behavior of the hitting probabilities for very large ν. This is achieved by

theorem 1. Put

$$\rho_k = f_{k+1} + f_{k+2} + \cdots \tag{10.13}$$

and recall from XI,(1.8) that the mean recurrence time satisfies

$$\mu = \rho_1 + \rho_2 + \cdots. \tag{10.14}$$

If $\mathcal{E}$ is not periodic we conclude from theorem 1 that as $\nu \to \infty$

$$w_\nu(r) \to \begin{cases} \rho_r/\mu & \textit{if} \quad \mu < \infty \\ 0 & \textit{if} \quad \mu = \infty. \end{cases} \tag{10.15}$$

This result is of great interest. In the case of a finite mean recurrence time it implies that $\{\rho_r/\mu\}$ is a probability distribution, and hence we have a limit theorem of a standard type. If, however, $\mu = \infty$ *the probability tends to* 1 *that the waiting time will exceed any given integer* r. In other words, our waiting times behave much worse than the recurrence times themselves. This unexpected phenomenon has significant consequences discussed in detail in volume 2. (See also problem 10.)

(*c*) *Repeated averaging.* The following problem is of an analytic character and was treated in various contexts by much more intricate methods. Suppose that $f_1 + \cdots + f_r = 1$ with $f_j \geq 0$. Given any r numbers $v_1, \ldots, v_r$ we define $f_1 v_r + \cdots + f_r v_1$ as their *weighted average*. We now define an infinite sequence $v_1, v_2, \ldots$ starting with the given r-tuple and defining v_n as the weighted average of the preceding r terms. In other words, for $n > r$ we define

$$v_n = f_1 v_{n-1} + \cdots + f_r v_{n-r}. \tag{10.16}$$

Since the sequence $f_1, f_2, \ldots$ terminates with the rth term these equations are of the form (10.1). We now define the b_k so that (10.1) will be true for all n. This means that we put $b_0 = v_0 = 0$ and

$$b_k = v_k - f_1 v_{k-1} - \cdots - f_{k-1} v_1 \qquad k \leq r. \tag{10.17}$$

(For $k > r$, by definition $b_k = 0$.) Without any calculations it follows from theorem 1 that with this repeated averaging the v_n *tend to a finite limit.* To calculate the limit we have to evaluate $b = b_1 + \cdots + b_r$. With the notation (10.13) for the remainders of $\sum f_k$ it is obvious from (10.17) and (10.6) that

$$v_n \to \frac{v_1 \rho_{r-1} + \cdots + v_r \rho_0}{f_1 + 2f_2 + \cdots + rf_r}. \tag{10.18}$$

For example, if $r = 3$ and one takes *arithmetic means*, then $f_1 = f_2 = = f_3 = \frac{1}{3}$ and

$$v_n \to \tfrac{1}{6}(v_1 + 2v_2 + 3v_3). \tag{10.19}$$

The ease with which we derived this result should not obscure the fact that the problem is difficult when taken out of the present context. (For an alternative treatment see problem 15 of XV,14.)

TABLE 1

ILLUSTRATING THE DEVELOPMENT OF THE AGE DISTRIBUTION IN A POPULATION DESCRIBED IN EXAMPLE (10.*d*)

n:	0	1	2	3	4	5	6	7	∞
$k = 0$	500	397	411.4	412	423.8	414.3	417.0	416.0	416.7
1	320	400	317.6	329.1	329.6	339.0	331.5	333.6	333.3
2	74	148	185	146.9	152.2	152.4	156.8	153.3	154.2
3	100	40	80	100	79.4	82.3	82.4	84.8	83.3
4	6	15	6	12	15	11.9	12.3	12.4	12.5

The columns give the age distribution of a population of $N = 1000$ elements at epochs $n = 0, 1, \ldots, 7$ together with the limiting distribution. The assumed mortalities are[10]

$$f_1 = 0.20; \quad f_2 = 0.43; \quad f_3 = 0.17; \quad f_4 = 0.17; \quad f_5 = 0.03,$$

so that no piece effectively attains age 5.

(*d*) *Self-renewing aggregates*. We return to the example of section 2 where a piece of equipment installed at epoch n has a lifetime with probability distribution $\{f_n\}$. When it expires it is immediately replaced by a new piece of the same character, and so the successive replacements constitute a persistent recurrent event in a sequence of dependent trials (whose outcomes decide whether or not a replacement takes place).

Suppose now that the piece of equipment installed at epoch 0 has an age k rather than being new. This affects only the first waiting time, and so $\mathcal{E}$ becomes a *delayed* recurrent event. To obtain the distribution $\{b_n\}$ of the first waiting time note that b_n is the (conditional) expectation that a piece will expire at age $n + k$ given that it has attained age k. Thus for $k \geq 1$

$$b_n = f_{n+k}/r_k \qquad \textit{where} \quad r_k = f_{k+1} + f_{k+2} + \cdots. \tag{10.20}$$

In practice one is not interested in a single piece of equipment but in a whole population (say the street lamps in a town). Suppose then that *the initial population* (at epoch 0) *consists of* N *pieces, among which* β_k *have age* k (where $\sum \beta_k = N$). Each piece originates a line of descendants which may require a replacement at epoch n. *The expected number* v_n *of all replacements at epoch* n *obviously satisfies the basic equations* (10.1) *with*

$$b_n = \sum \beta_k f_{n+k}/r_k. \tag{10.21}$$

We have here the first example where v_n is an *expectation* rather than a probability; we know only that $v_n < N$.

An easy calculation shows that $b = \sum b_n = N$, and so theorem 1 shows that $v_n \to N/\mu$ provided that the replacements are not periodic. This result implies the existence of a *stable limit for the age distribution*. In fact, for a piece to be of age k at epoch n it is necessary and sufficient that it was installed at epoch $n - k$ and that it survived age k. The expected number of such pieces is therefore $v_{n-k}r_k$ and tends to Nr_k/μ as $n \to \infty$. In other words, as time goes on the *fraction of the population of age* k *tends to* r_k/μ. Thus *the limiting age distribution is independent of the initial age distribution* and depends only on the mortalities f_n. A similar result holds under much wider conditions. For a numerical illustration see table 1. It reveals the noteworthy fact that the approach to the limit is not monotone. (See also problems 16–18.)

(*e*) *Human populations*. For an example where $f = \sum f_n > 1$ we use the simplest model of a human population. It is analogous to the model in the preceding example except that the population size is now variable and female births take over the role of replacements. The novel feature is that a mother may have any number of daughters, and hence her line may become extinct, but it may also increase in numbers. We now define f_n the probability, at birth, that a mother will (survive and) at age n give birth to a female child. (The dependence on the number and the ages of previous children is neglected.) Then $f = \sum f_n$ is the expected number of daughters and so in a healthy population $f > 1$. Theorem 1 then promises a population size that increases roughly at the constant rate ξ, and the age distribution of the population tends to a limit as described in the preceding example. The model is admittedly crude but presents nevertheless some practical interest. The curious dependence of the limiting behavior ξ was certainly not predictable without a proper mathematical analysis. ▶

[10] The roots of the equation $1 - F(s) = 0$ are $1, -\frac{5}{3}, -5$, and $\pm 2i$. The mean recurrence time is 2.40.

*11. PROOF OF THE BASIC LIMIT THEOREM

In section 3 we omitted the proof of theorem 3 which we now restate as follows: *Let* $f_1, f_2, \ldots$ *be a sequence of numbers* $f_n \geq 0$ *such that*

[*] This section is not used in the sequel.

$\sum f_n = 1$ *and* 1 *is the greatest common divisor of those* n *for which* $f_n > 0$. *Let* $u_0 = 1$ *and*

$$u_n = f_1 u_{n-1} + f_2 u_{n-2} + \cdots + f_n u_0, \qquad n \geq 1. \tag{11.1}$$

Then

$$u_n \to \mu^{-1} \qquad \textit{where} \quad \mu = \sum_{n=1}^{\infty} n f_n \tag{11.2}$$

(μ^{-1} *being interpreted as* 0 *when* $\mu = \infty$).

In order not to interrupt the argument we preface the proof by two well-known lemmas that are widely used outside probability.

Let A be the set of all integers n for which $f_n > 0$, and denote by A^+ the set of all positive linear combinations

$$p_1 a_1 + \cdots + p_r a_r \tag{11.3}$$

of numbers $a_1, \ldots, a_r$ in A (the p_j are positive integers).

Lemma 1. *There exists an integer* N *such that* A^+ *contains all integers* $n > N$.

Proof. As is known from Euclid, the fact that 1 is the greatest common divisor of the numbers in A means that it is possible to choose integers $a_1, \ldots, a_r$ in A and (not necessarily positive) integers c_j such that

$$c_1 a_1 + \cdots + c_r a_r = 1. \tag{11.4}$$

Put $s = a_1 + \cdots + a_r$. *Every* integer n admits of a unique representation $n = xs + y$ where x and y are integers and $0 \leq y < s$. Then

$$n = \sum_{k=1}^{r} (x + c_k y) a_k \tag{11.5}$$

and all the coefficients will be positive as soon as x exceeds y times the largest among the numbers $|c_k|$. ▶

Lemma 2. (*Selection principle.*) *Suppose that for every integer* $\nu > 0$ *we are given a sequence of numbers* $z_1^{(\nu)}, z_2^{(\nu)}, \ldots$ *such that* $0 \leq z_k^{(\nu)} \leq 1$. *Then there exists a sequence* $\nu^{(1)}, \nu^{(2)}, \ldots \to \infty$ *such that as* ν *runs through it,* $z_k^{(\nu)}$ *tends to a limit for every fixed* k.

Proof[11] Choose an increasing sequence $\nu_1^{(1)}, \nu_2^{(1)}, \ldots$ such that as ν runs through it $z_1^{(\nu)}$ converges to a limit z_1. Out of this sequence choose

[11] The proof is based on the so-called *diagonal method* due to G. Cantor (1845–1918). It has become a standard tool but was shockingly new in Cantor's time.

a *subsequence* $\nu_1^{(2)}, \nu_2^{(2)}, \ldots$ such that as ν runs through it $z_2^{(\nu)} \to z_2$. Continuing in this way we get for each n a sequence of integers $\nu_j^{(n)} \to \infty$ such that as ν runs through it $z_n^{(\nu)} \to z_n$, and each $\nu_j^{(n)}$ is an element of the preceding sequence $\{\nu_j^{(n-1)}\}$. Finally, put $\nu^{(r)} = \nu_r^{(r)}$. Let $r > n$. Except for the first n terms every element $\nu^{(r)}$ appears in the sequence $\nu_1^{(n)}, \nu_2^{(n)}, \ldots,$ and hence $z_n^{(\nu)} \to z_n$ as ν runs through the sequence $\nu^{(1)}, \nu^{(2)}, \ldots.$ ▶

Lemma 3. *Let* $\{w_n\}$ $(n = 0, \pm 1, \pm 2, \ldots)$ *be a doubly infinite sequence of numbers such that* $0 \leq w_n \leq 1$ *and*

$$w_n = \sum_{k=1}^{\infty} f_k w_{n-k} \tag{11.6}$$

for each n. *If* $w_0 = 1$ *then* $w_n = 1$ *for all* n.

Proof. Since

$$w_0 = \sum_{k=1}^{\infty} f_k w_{-k} \leq \sum_{k=1}^{\infty} f_k = 1 \tag{11.7}$$

the condition $w_0 = 1$ requires that the two series agree termwise, and so for each k either $f_k = 0$ or else $w_{-k} = 1$. This means that $w_{-a} = 1$ for every integer a of A. But then the argument used for $n = 0$ applies also with $n = -a$, and we conclude that $w_{-a-b} = 1$ whenever the integers a and b are in A. Proceeding by induction we conclude that $w_{-m} = 1$ for every integer in A^+, and hence $w_{-m} = 1$ for every $m > N$. But this implies that for $n = -N$ the right side in (11.6) equals 1 and so $w_{-N} = 1$. Letting $n = -N + 1$ we find in like manner $w_{-N+1} = 1$, and proceeding in this way we find by induction that $w_n = 1$ for all n. ▶

Proof of the theorem. Let

$$\eta = \limsup_{n \to \infty} u_n. \tag{11.8}$$

It is obvious from (11.1) that $0 \leq \eta \leq 1$, and there exists a sequence $r_1, r_2, \ldots$ tending to infinity such that as $\nu \to \infty$

$$u_{r_\nu} \to \eta. \tag{11.9}$$

For each positive integer ν we define a doubly infinite sequence $\{u_n^{(\nu)}\}$ by

$$u_n^{(\nu)} = \begin{array}{ll} u_{r_\nu + n} & \textit{for} \quad n \geq -r_\nu \\ 0 & \textit{for} \quad n < -r_\nu. \end{array} \tag{11.10}$$

For simplicity of expression lemma 2 was formulated for simple sequences, but it obviously applies to double sequences also. Accordingly, it is possible to choose an increasing sequence of integers $\nu_1, \nu_2, \ldots$ such that when ν runs through it $u_n^{(\nu)}$ tends to a limit w_n for each n. From the construction $0 \leq w_n \leq \eta$ and $w_0 = \eta$. Furthermore, for each ν and $n > -\nu$ the definition (11.1) reads

$$u_n^{(\nu)} = \sum_{k=1}^{\infty} f_k u_{n-k}^{(\nu)}, \tag{11.11}$$

and in the limit we find the relation (11.6). By lemma 3 therefore $w_n = \eta$ for all n.

We are now ready for the final argument. As before we put

$$\rho_k = f_{k+1} + f_{k+2} + \cdots \tag{11.12}$$

so that $r_0 = 1$ and $\sum \rho_k = \mu$ [see XI, (1.8)]. Summing the defining relations (11.1) over $n = 1, 2, \ldots, N$ and collecting terms we get the identity

$$\rho_0 u_N + \rho_1 u_{N-1} + \cdots + \rho_N u_0 = 1. \tag{11.13}$$

We use this relation successively for $N = \nu_1, \nu_2, \ldots$. As N runs through this sequence $u_{N-k} \rightarrow w_{-k} = \eta$ for each k. If $\sum \rho_k = \infty$ it follows that $\eta = 0$ and so $u_n \rightarrow 0$ as asserted. When $\mu = \sum \rho_k < \infty$ it follows that $\eta = \mu^{-1}$, and it remains to show that this implies $u_N \rightarrow \eta$ for any approach $N \rightarrow \infty$. By the definition of the upper limit we have $u_{N-k} < < \eta + \epsilon$ for each fixed k and N sufficiently large. Furthermore $u_n \leq 1$ for all n. Suppose then that N approaches infinity in such a manner that $u_N \rightarrow \eta_0$. From (11.13) it is clear that ultimately

$$\rho_0\eta_0 + (\rho_1 + \cdots + \rho_r)(\eta + \epsilon) + (\rho_{r+1} + \rho_{r+2} + \cdots) \geq 1, \tag{11.14}$$

and hence

$$\rho_0(\eta_0 - \eta) + \mu(\eta + \epsilon) \geq 1. \tag{11.15}$$

But $\mu\eta = 1$ and $\eta_0 \leq \eta$ by the definition of η. Since (11.15) is true for arbitrary $\epsilon > 0$ it follows that $\eta_0 = \eta$ and so $u_N \rightarrow \mu^{-1}$ for any approach $N \rightarrow \infty$. ►

12. PROBLEMS FOR SOLUTION

1. Suppose that $F(s)$ is a polynomial. Prove for this case all theorems of section 3, using the partial fraction method of XI, 4.

2. Let r coins be tossed repeatedly and let $\mathcal{E}$ be the recurrent event that for each of the r coins the accumulated number of heads and tails are equal. Is $\mathcal{E}$ persistent or transient? For the smallest r for which $\mathcal{E}$ is transient, estimate the probability that $\mathcal{E}$ ever occurs.

3. In a sequence of independent throws of a perfect die let $\mathcal{E}$ stand for the event that the accumulated numbers of ones, twos, ..., sixes are equal. Show that $\mathcal{E}$ is a transient (periodic) recurrent event and estimate the probability f that $\mathcal{E}$ will ever occur.

4. In a sequence of Bernoulli trials let $\mathcal{E}$ occur when the accumulated number of successes equals λ times the accumulated number of failures; here λ is a positive integer. [See example (1.*c*).] Show that $\mathcal{E}$ is persistent if, and only if, $p/q = \lambda$, that is, $p = \lambda/(\lambda+1)$. *Hint:* Use the normal approximation.

5. In a sequence of Bernoulli trials we say that $\mathcal{E}$ occurs when the accumulated number of successes is twice the accumulated number of failures and the ratio has never exceeded 2. Show that $\mathcal{E}$ is transient and periodic. The generating function is determined by the cubic equation $F(s) = qs(U(s)ps)^2$. (*Hint:* $U(s)ps$ is the generating function for the waiting time for the number of successes to *exceed* twice the number of failures.)

6. Let the $\mathbf{X}_j$ be independent integral-valued random variables with a common distribution. Assume that these variables assume both positive and negative values. Prove that the event defined by $\mathbf{S}_n = 0, \mathbf{S}_1 \leq 0, \ldots, \mathbf{S}_{n-1} \leq 0$ is recurrent and transient.

7. *Geiger counters.* [See examples (1.*g*) and (4.*e*).] Denote by $\mathbf{N}_n$ and $\mathbf{Z}_n$, respectively, the number of occurrences of $\mathcal{E}$ and the number of registrations up to and including epoch n. Discuss the relationship between these variables and find asymptotic expressions for $\mathbf{E}(\mathbf{Z}_n)$ and $\mathrm{Var}(\mathbf{Z}_n)$.

8. In *Geiger counters of type* II every arriving particle (whether registered or not) locks the counter for exactly r time units (that is, at the $r - 1$ trials following the arrival). The duration of the locked time following a registration is therefore a random variable. Find its generating function G. If $\mathcal{E}$ is again the recurrent event that the counter is free, express the generating function F of the recurrence times in terms of G. Finally, find the mean recurrence time.

9. *A more general type of Geiger counters.* As in problem 8 we assume that every arriving particle completely obliterates the effect of the preceding ones, but we assume now that the time for which a particle locks the counter is a random variable with a given generating function $B(s)$. [In the preceding problem $B(s) = s^r$.] Do problem 8 under these more general conditions.

10. For a delayed recurrent event $\mathcal{E}$ the probabilities v_n are constant only when the generating function of the first occurrence of $\mathcal{E}$ is given by $B(s) = = [1 - F(s)]/\mu(1-s)$, that is, when $b_n = f_{n+1} + f_{n+2} + \cdots$. Discuss the relation with the limit theorem for hitting probabilities in example (10.*b*).

11. Find an approximation to the probability that in 10,000 tossings of a coin the number of head runs of length 3 will lie between 700 and 730.

12. In a sequence of tossings of a coin let $\mathcal{E}$ stand for the pattern *HTH*. Let r_n be the probability that $\mathcal{E}$ does not occur in n trials. Find the generating function and use the partial fraction method to obtain an asymptotic expansion.

13. In example (8.*a*) the expected duration of the game is

$$\mu_1\mu_2/(\mu_1 + \mu_2),$$

where μ_1 and μ_2 are the mean recurrence times for success runs of length r and failure runs of length ρ, respectively.

14. The possible outcomes of each trial are A, B, and C; the corresponding probabilities are α, β, γ $(\alpha + \beta + \gamma = 1)$. Find the generating function of the probability that in n trials there is no run of length r: (*a*) of A's, (*b*) of A's or B's, (*c*) of any kind.

15. *Continuation.* Find the probability that the first A-run of length r precedes the first B-run of length ρ and terminates at the nth trial. (*Hint:* The generating function is of the form $\mathfrak{X}(s)$ in (8.6) except that ρ is replaced by α in the expression for F, and by β in $\mathfrak{G}$.)

16. *Self-renewing aggregates.* In example (10.*d*) find the limiting age distribution assuming that the lifetime distribution is geometric: $f_k = q^{k-1}p$.

17. *Continuation.* The initial *age distribution* $\{\beta_k\}$ *is called stationary* if it perpetuates itself for all times. Show (without computation) that this is the case only when $\beta_k = r_k/\mu$.

18. *Continuation.* Denote by $w_k(n)$ the expected number of elements at epoch n that are of age k. Find the determining equations and verify from them that the population size remains constant. Furthermore, show that the expected number $w_0(n)$ satisfies

$$w_0(n) = w_0(n-1)f_1/r_0 + w_1(n-1)f_2/r_1 + \cdots.$$

19. Let $\mathcal{E}$ be a persistent aperiodic recurrent event. Assume that the recurrence time has finite mean μ and variance σ^2. Put $q_n = f_{n+1} + f_{n+2} + \cdots$ and $r_n = q_{n+1} + q_{n+2} + \cdots$. Show that the generating functions $Q(s)$ and $R(s)$ converge for $s = 1$. Prove that

$$u_0 + \sum_{n=1}^{\infty}\left(u_n - \frac{1}{\mu}\right)s^n = \frac{R(s)}{\mu Q(s)} \tag{12.1}$$

and hence that

$$u_0 + \sum_{n=1}^{\infty}\left(u_n - \frac{1}{\mu}\right) = \frac{\sigma^2 - \mu + \mu^2}{2\mu^2}. \tag{12.2}$$

20. Let $\mathcal{E}$ be a persistent recurrent event and $\mathbf{N}_r$ the number of occurrences of $\mathcal{E}$ in r trials. Prove that

$$\mathbf{E}(\mathbf{N}_r^2) = u_1 + \cdots + u_r + 2\sum_{j=1}^{r-1} u_j(u_1 + \cdots + u_{r-j}) \tag{12.3}$$

and hence that $\mathbf{E}(\mathbf{N}_r^2)$ is the coefficient of s^r in

$$\frac{F^2(s) + F(s)}{(1-s)\{1-F(s)\}^2} \tag{12.4}$$

(Note that this may be reformulated more elegantly using bivariate generating functions.)

21. Let $q_{k,n} = \mathbf{P}\{\mathbf{N}_k = n\}$. Show that $q_{k,n}$ is the coefficient of s^k in

$$F^n(s)\frac{\{1-F(s)\}}{1-s}. \tag{12.5}$$

Deduce that $\mathbf{E}(\mathbf{N}_r)$ and $\mathbf{E}(\mathbf{N}_r^2)$ are the coefficients of s^r in

$$\frac{F(s)}{(1-s)\{1-F(s)\}} \tag{12.6}$$

and (12.4), respectively.

22. Using the notations of problem 19, show that

$$\frac{F(s)}{(1-s)\{1-F(s)\}} = -\frac{1}{1-s} + \frac{1}{\mu(1-s)^2} + \frac{R(s)}{\mu\{1-F(s)\}}. \tag{12.7}$$

Hence, using the last problem, conclude that

$$\mathbf{E}(\mathbf{N}_r) = \frac{r}{\mu} + \frac{\sigma^2 + \mu - \mu^2}{2\mu^2} + \epsilon_r \tag{12.8}$$

with $\epsilon_r \to 0$.

23. *Continuation.* Using a similar argument, show that

$$\mathbf{E}(\mathbf{N}_r^2) = \frac{r^2}{\mu^2} + \frac{2\sigma^2 + \mu - \mu^2}{\mu^3}r + \alpha_r, \tag{12.9}$$

where $\alpha_r/r \to 0$. Hence

$$\text{Var}(\mathbf{N}_r) \sim \frac{\sigma^2}{\mu^3}r. \tag{12.10}$$

(*Hint:* Decompose the difference of (12.4) and (12.7) into three fractions with denominators containing the factor $(1-s)^k$, $k = 1, 2, 3$.)

24. In a sequence of Bernoulli trials let $q_{k,n}$ be the probability that exactly n success runs of length r occur in k trials. Using problem 21, show that the generating function $Q_k(x) = \sum q_{k,n}x^n$ is the coefficient of s^k in

$$\frac{1 - p^r s^r}{1 - s + qp^r s^{r+1} - (1-ps)p^r s^r x}.$$

Show, furthermore, that the root of the denominator which is smallest in absolute value is $s_1 \approx 1 + qp^r(1-x)$.

25. *Continuation. The Poisson distribution of long runs.*[12] If the number k of trials and the length r of runs both tend to infinity, so that $kqp^r \to \lambda$, then the probability of having exactly n runs of length r tends to $e^{-\lambda}\lambda^n/n!$.

Hint: Using the preceding problem, show that the generating function is asymptotically $\{1 + qp^r(1-x)\}^{-k} \sim e^{-\lambda(1-x)}$. Use the *continuity theorem* of XI, 6.

[12] The theorem was proved by von Mises, but the present method is considerably simpler.