THE THEORY OF BRANCHING PROCESSES

VON

THEODORE E. HARRIS

THE THEORY OF BRANCHING PROCESSES

BY

THEODORE E. HARRIS
DEPARTMENT OF MATHEMATICS,
THE RAND CORPORATION, SANTA MONICA, CALIFORNIA

WITH 6 FIGURES

SPRINGER-VERLAG
BERLIN · GÖTTINGEN · HEIDELBERG
1963

DIE GRUNDLEHREN DER

MATHEMATISCHEN WISSENSCHAFTEN

IN EINZELDARSTELLUNGEN MIT BESONDERER BERÜCKSICHTIGUNG DER ANWENDUNGSGEBIETE

BAND 119

SPRINGER-VERLAG
BERLIN · GÖTTINGEN · HEIDELBERG
1963

Preface

It was about ninety years ago that GALTON and WATSON, in treating the problem of the extinction of family names, showed how probability theory could be applied to study the effects of chance on the development of families or populations. They formulated a mathematical model, which was neglected for many years after their original work, but was studied again in isolated papers in the twenties and thirties of this century.

During the past fifteen or twenty years, the model and its generalizations have been treated extensively, for their mathematical interest and as a theoretical basis for studies of populations of such objects as genes, neutrons, or cosmic rays. The generalizations of the Galton-Watson model to be studied in this book can appropriately be called *branching processes*; the term has become common since its use in a more restricted sense in a paper by KOLMOGOROV and DMITRIEV in 1947 (see Chapter II). We may think of a branching process as a mathematical representation of the development of a population whose members reproduce and die, subject to laws of chance. The objects may be of different types, depending on their age, energy, position, or other factors. However, they must not interfere with one another. This assumption, which unifies the mathematical theory, seems justified for some populations of physical particles such as neutrons or cosmic rays, but only under very restricted circumstances for biological populations.

Chapter I studies the original model of GALTON and WATSON, which was designed to answer the following question: If a man has probabilities $p_0, p_1, p_2, \ldots$ for having 0, 1, 2, ... sons, if each son has the same probabilities for sons of his own, and so on, what is the probability that the family will eventually become extinct, and more generally, what is the probability of a given number of male descendants in a given generation? Chapter II deals with a natural generalization, where each object may be one of several types, and Chapter III carries on the generalization, so that one can deal with objects described by continuous variables such as age, energy, etc. The theory is then applied in Chapter IV to some of the simpler mathematical models for neutron chain reactions. Chapter V treats the model of GALTON and WATSON in cases where the development of a family is traced continuously in time, rather than by generations, and Chapter VI describes the most natural way of treating populations whose objects are subject to aging effects. Finally, Chapter VII describes a mathematical theory of the electron-photon cascade, one of the components of cosmic radiation.

In this book, the emphasis is on a systematic development of the mathematical theory, but I have described briefly the more important applications, indicating in a general way the weak points of the assumptions underlying some of them. The mathematical level of the treatment varies. I believe that most of Chapters I, II, and V can be mastered by anyone with a working knowledge of Markov chains and continuous probability distributions, at the level of FELLER'S *Probability Theory* and PARZEN'S *Modern Probability Theory*, respectively. I hope that such readers can at least follow the main results in the remainder of the book, whose detailed reading requires an amount of measure-theoretic probability about equal to that in KOLMOGOROV'S basic monograph *Foundations of Probability Theory*, plus a few results from the more advanced treatises of DOOB and LOÈVE. Occasional use is made of matrix theory, the theory of analytic functions, and the theory of Fourier and Laplace integrals.

Although I have tried to give rigorous proofs of the more basic results, I have not hesitated to include a heuristic proof (so labeled) when I did not know a rigorous one, or when the length of a rigorous one seemed out of proportion to its importance.

Thanks are due D. A. DARLING and RUPERT MILLER, who read the entire manuscript and suggested numerous improvements, and also my colleague RICHARD BELLMAN, who made many suggestions about the presentation. I wish to thank my former teacher, S. S. WILKS, who introduced me to this problem, and J. L. DOOB, who encouraged me to write the book. I appreciate the excellent work of MARGARET WRAY, who typed several versions of the manuscript, and of ELEANOR HARRIS, who prepared it for the printer.

I am indebted to The RAND Corporation for support of this work under a broad research contract with the United States Air Force.

Finally, I want to thank my wife for her patience during the many evenings when I was busy with this book.

Santa Monica, California — T. E. H.

July, 1963

Contents

Chapter I

The Galton-Watson branching process

1. Historical remarks

The decay of the families of men who occupied conspicuous positions in past times has been a subject of frequent remark, and has given rise to various conjectures ... The instances are very numerous in which surnames that were once common have since become scarce or have wholly disappeared. The tendency is universal, and, in explanation of it, the conclusion has been hastily drawn that a rise in physical comfort and intellectual capacity is necessarily accompanied by diminution in "fertility" ... On the other hand, M. ALPHONSE DE CANDOLLE has directed attention to the fact that, by the ordinary law of chances, a large proportion of families are continually dying out, and it evidently follows that, until we know what that proportion is, we cannot estimate whether any observed diminution of surnames among the families whose history we can trace, is or is not a sign of their diminished "fertility".

These remarks of FRANCIS GALTON were prefaced to the solution by the Reverend H. W. WATSON of the "problem of the extinction of families", which appeared in 1874[1]. Not willing to accept uncritically the hypothesis that distinguished families are more likely to die out than ordinary ones, GALTON recognized that a first step in studying the hypothesis would be to determine the probability that an ordinary family will disappear, using fertility data for the whole population. Accordingly, he formulated the problem of the extinction of families as follows:

Let $p_0, p_1, p_2, \ldots$ be the respective probabilities that a man has 0, 1, 2, ... sons, let each son have the same probability for sons of his own, and so on. What is the probability that the male line is extinct after r generations, and more generally what is the probability for any given number of descendants in the male line in any given generation?

WATSON's ingenious solution of this problem used a device that has been basic in most subsequent treatments. However, because of a purely algebraic oversight, WATSON concluded erroneously that every family will die out, even when the population size, on the average, increases from one generation to the next.

Although we shall not be concerned with questions of demography, let us note at this point that GALTON (1891) studied statistics on the reproductive rates of English peers, coming to the interesting conclusion

[1] WATSON and GALTON (1874). Essentially the same discussion was given in an appendix to GALTON's book (1889). GALTON originally posed the problem in the pages of the *Educational Times*.

that one factor in lowering the rates was the tendency of peers to marry heiresses. An heiress, coming from a family with no sons, would be expected to have, by inheritance, a lower-than-ordinary fertility, and GALTON's data bore out this expectation.

The mathematical model of GALTON and WATSON (we shall call it the *Galton-Watson process*) appears to have been neglected for many years after its creation, the next treatment known to the author being that of R. A. FISHER (1922, 1930a, 1930b). FISHER used a mathematical model identical with that of GALTON and WATSON to study the survival of the progeny of a mutant gene and to study random variations in the frequencies of genes. J. B. S. HALDANE (1927) likewise applied the model to genetics.

The first complete and correct determination of the probability of extinction for the Galton-Watson process was given by J. F. STEFFENSEN (1930, 1932). The problem was also treated by A. KOLMOGOROV (1938), who determined the asymptotic form of the probability that the family is still in existence after a large finite number of generations.

A. J. LOTKA (1931a, 1931b, 1939a) carried out GALTON's idea, using American fertility data, to determine the probability of extinction of a male line of descent. N. SEMENOFF (1935, Chapter III) used the Galton-Watson model in the elementary stages of his theoretical treatise on chemical (as opposed to nuclear) chain reactions, and W. SHOCKLEY and J. R. PIERCE (1938) employed the model to study the multiplication of electrons in an electronic detection device (the electron multiplier).

After 1940 interest in the model increased, partly because of the analogy between the growth of families and nuclear chain reactions, and partly because of the increased general interest in applications of probability theory. Early work stimulated by the nuclear analogy included that of D. HAWKINS and S. ULAM (1944) and C. J. EVERETT and S. ULAM (1948a, b, c, d). During the past 15 years the model has been the subject of numerous papers in Britain, the Soviet Union, and the United States.

The original Galton-Watson process and its generalizations are connected with work dating back to NIELS ABEL[1] on functional equations and the iteration of functions, and with various lines of development in the theory of stochastic processes. For example, there is an interesting connection between the Galton-Watson process and the so-called birth-and-death processes, introduced in a special form by G. U. YULE (1924) in a study of the rate of formation of new species. The species, rather than individual animals, are the multiplying objects. There are also connections with the theory of cosmic radiation formulated independently

[1] ABEL (1881). This posthumous paper appears in ABEL's collected works.

by H. J. BHABHA and W. HEITLER (1937) and by J. F. CARLSON and J. R. OPPENHEIMER (1937).

These biological and physical problems have required treatment by mathematical models more elaborate than the Galton-Watson process, which is the subject of the present chapter. Although some of these later models are only remotely related to the Galton-Watson process, others can justifiably be considered its direct descendants. It is with these that the chapters after the first will be principally concerned.

With few exceptions we shall treat only processes in which it is assumed that different objects reproduce *independently* of one another. This is a severe limitation for any application to biological problems, although there are situations, which we shall point out, where the assumption of independence seems reasonable. For many processes of interest in physics the assumption of independence seems realistic, although, of course, the models are always imperfect in other ways.

2. Definition of the Galton-Watson process

Let us imagine objects that can generate additional objects of the same kind; they may be men or bacteria reproducing by familiar biological methods, or neutrons in a chain reaction. An initial set of objects, which we call the 0-th generation, have children that are called the first generation; their children are the second generation, and so on. The process is affected by chance events.

In this chapter we choose the simplest possible mathematical description of such a situation, corresponding to the model of GALTON and WATSON. First of all we keep track only of the sizes of the successive generations, not the times at which individual objects are born or their individual family relationships. We denote by $Z_0, Z_1, Z_2, \ldots$ the numbers in the 0-th, first, second, ... generations. (We can sometimes interpret $Z_0, Z_1, \ldots$ as the sizes of a population at a sequence of points in time; see Secs. V.5 and VI.27.) Furthermore, we make the two following assumptions.

(1) If the size of the n-th generation is known, then the probability law governing later generations does not depend on the sizes of generations preceding the n-th; in other words, $Z_0, Z_1, \ldots$ form a *Markov chain*. We shall nearly always make the additional assumption that the transition probabilities for the chain do not vary with time.

(2) The Markov chains considered in this chapter have a very special property, corresponding to the assumption that different objects do not interfere with one another: The number of children born to an object does not depend on how many other objects are present.

Assumption (1) could fail, for example, if a man with few brothers tends to have fewer sons than a man with many brothers. In this case it would help us to know whether a generation comprising six men were all brothers or came from three different fathers. We might restore the Markovian nature of the mathematical model by introducing different types corresponding to different fertilities. This would lead to the models of Chapters II and III.

Assumption (2) fails if the different objects interact with one another. Some discussion of this point for biological populations is given in Secs. 7.3, V.2, and VI.23. The assumption is supposed to be good for particles such as those of the neutron processes of Chapter IV and the electron-photon cascades of Chapter VII.

The author will occasionally remind the reader of the weak points in applications of the various mathematical models to be introduced. However, there will be no systematic attempt to evaluate the worth of the various assumptions.

2.1. Mathematical description of the Galton-Watson process. Let $Z_0, Z_1, Z_2, \ldots$ denote the successive random variables in our Markov process (more particularly, Markov chain, since the states in the process are nonnegative integers). We interpret Z_n as the number of objects in the n-th generation of a population or family. *We shall always assume that $Z_0=1$, unless the contrary is stated.* The appropriate adjustments if $Z_0 \neq 1$ are easily made, because we assume that the families of the initial objects develop independently of one another.

We denote by P the probability measure for our process. The probability distribution of Z_1 is prescribed by putting $P(Z_1=k)=p_k$, $k=0, 1, 2, \ldots$, $\Sigma p_k=1$, where p_k is interpreted as the probability that an object existing in the n-th generation has k children in the $(n+1)$-th generation. It is assumed that p_k does not depend on the generation number n.

The conditional distribution of Z_{n+1}, given $Z_n=k$, is appropriate to the assumption that different objects reproduce independently; that is, Z_{n+1} is distributed as the sum of k independent random variables, each distributed like Z_1. If $Z_n=0$, then Z_{n+1} has probability 1 of being 0. Thus we have defined the *transition probabilities* of our Markov process, denoted by

$$P_{ij}=P(Z_{n+1}=j|Z_n=i), \qquad i, j, n=0, 1, \ldots . \tag{2.1}$$

These transition probabilities are defined for each i and j even though, strictly speaking, the right side of (2.1) is not defined as a conditional probability if $P(Z_n=i)=0$.

Having defined the process, we shall want to know some of its properties: the probability distribution and moments of Z_n; the probability that the random sequence $Z_0, Z_1, Z_2, \ldots$ eventually goes to zero; and the behavior of the sequence in case it does not go to zero.

2.2. Generating functions. We shall make repeated use of the *probability generating function*

$$f(s)=\sum_{k=0}^{\infty} p_k s^k, \qquad |s|\leqq 1, \tag{2.2}$$

where s is a complex variable.

The *iterates* of the generating function $f(s)$ will be defined by

$$f_0(s)=s, \qquad f_1(s)=f(s), \tag{2.3}$$

$$f_{n+1}(s)=f[f_n(s)], \qquad n=1, 2, \ldots. \tag{2.4}$$

The reader can verify that each of the iterates is a probability generating function, and that the following relations are a consequence of (2.3) and (2.4):

$$f_{m+n}(s)=f_m[f_n(s)], \qquad m, n=0, 1, \ldots, \tag{2.5}$$

and in particular,

$$f_{n+1}(s)=f_n[f(s)]. \tag{2.6}$$

3. Basic assumptions

Throughout this chapter we shall, without further mention, make the following assumptions, unless the contrary is stated.

(a) None of the probabilities $p_0, p_1, \ldots$ is equal to 1, and $p_0+p_1<1$. Thus f is strictly convex on the unit interval.

(b) The expected value $\mathscr{E}Z_1=\sum_{k=0}^{\infty} kp_k$ is finite. This implies that the derivative $f'(1)$ is finite. The symbols $f'(1)$, $f''(1)$, etc., will usually refer to left-hand derivatives at $s=1$, since we usually suppose $|s|\leqq 1$.

4. The generating function of Z_n

The following basic result was discovered by WATSON (1874) and has been rediscovered a number of times since. The Basic Assumptions are not required for this result.

Theorem 4.1. *The generating function of Z_n is the n-th iterate $f_n(s)$.*

Proof. Let $f_{(n)}(s)$ designate the generating function of Z_n, $n=0, 1, \ldots$. Under the condition that $Z_n=k$, the distribution of Z_{n+1} has the generating function $[f(s)]^k$, $k=0, 1, \ldots$. Accordingly the generating function of Z_{n+1} is

$$f_{(n+1)}(s)=\sum_{k=0}^{\infty} P(Z_n=k)[f(s)]^k=f_{(n)}[f(s)], \qquad n=0, 1, \ldots. \tag{4.1}$$

From the definitions of $f_{(0)}$ and f_0, we see that they are equal. Using (2.6) and (4.1), we then see by induction that $f_{(n)}(s)=f_n(s)$, $n=1, 2, \ldots$. □

Theorem 4.1 enables us to calculate the generating function, and hence the probability distribution, of Z_n in a routine manner by simply computing the iterates of f, although only rarely can the n-th iterate be found in a simple explicit form. From the point of view of probability theory, the main value of Theorem 4.1 is that it enables us to calculate the moments of Z_n and to obtain various asymptotic laws of behavior for Z_n when n is large.

Remark. We leave it to the reader to show that if k is a positive integer, then $Z_0, Z_k, Z_{2k}, Z_{3k}, \ldots$ is a Galton-Watson process with the generating function $f_k(s)$.

We now consider the moments of Z_n.

5. Moments of Z_n

Definitions 5.1. Let

$$m = \mathscr{E} Z_1, \qquad \sigma^2 = \text{Variance } Z_1 = \mathscr{E} Z_1^2 - m^2.$$

Note that $m = f'(1)$ and $\sigma^2 = f''(1) + m - m^2$.

We can obtain the moments of Z_n by differentiating either (2.4) or (2.6) at $s=1$. Thus differentiating (2.4) yields

$$f'_{n+1}(1) = f'[f_n(1)] f'_n(1) = f'(1) f'_n(1), \tag{5.1}$$

whence by induction $f'_n(1) = m^n$, $n = 0, 1, \ldots$. If $f''(1) < \infty$, we can differentiate (2.4) again, obtaining

$$f''_{n+1}(1) = f'(1) f''_n(1) + f''(1) [f'_n(1)]^2. \tag{5.2}$$

We obtain $f''_n(1)$ by repeated application of (5.2) with $n = 0, 1, 2, \ldots$; thus

$$\left.\begin{aligned} \text{Variance } Z_n = \mathscr{E} Z_n^2 - (\mathscr{E} Z_n)^2 &= \frac{\sigma^2 m^n (m^n - 1)}{m^2 - m}, \qquad m \neq 1; \\ &= n\sigma^2, \qquad m = 1. \end{aligned}\right\} \tag{5.3}$$

We thus have the following results[1].

Theorem 5.1. *The expected value $\mathscr{E} Z_n$ is m^n, $n = 0, 1, \ldots$. If $\sigma^2 =$ Variance $Z_1 < \infty$, then the variance of Z_n is given by* (5.3).

If higher moments of Z_1 exist, then higher moments of Z_n can be found in a similar fashion.

[1] STEFFENSEN (1932) showed how to obtain the moments in essentially this manner.