
4

BRANCHING PROCESSES

Every moment dies a man, every moment one is born.

—Alfred Tennyson, *The Vision of Sin*

Every moment dies a man, every moment one and one-sixteenth is born.

—Mathematician Charles Babbage in a letter to Alfred Tennyson
suggesting a change “in your otherwise beautiful poem.”

4.1 INTRODUCTION

Branching processes are a class of stochastic processes that model the growth of populations. They are widely used in biology and epidemiology to study the spread of infectious diseases and epidemics. Applications include nuclear chain reactions and the spread of computer software viruses. Their original motivation was to study the extinction of family surnames, an issue of concern to the Victorian aristocracy in 19th century Britain.

In 1873, the British statistician Sir Francois Galton posed the following question in the *Educational Times*.

Problem 4001: A large nation, of whom we will only concern ourselves with adult males, N in number, and who each bear separate surnames colonize a district. Their law of population is such that, in each generation, a_0 percent of the adult males have no male children who reach adult life; a_1 have one such male child; a_2 have two; and so on up to

a_5 who have five. Find (1) what proportion of their surnames will have become extinct after r generations; and (2) how many instances there will be of the surname being held by m persons.

The Reverend Henry William Watson replied with a solution. The study of branching processes grew out of Watson and Galton's collaboration. Their results were independently discovered by the French statistician Irénée-Jules Bienaymé. The basic branching process model is sometimes called a Bienaymé–Galton–Watson process.

We use the imagery of populations, generations, children, and offspring. Assume that we have a population of individuals, each of which independently produces a random number of children according to a probability distribution $\mathbf{a} = (a_0, a_1, a_2, \dots)$. That is, an individual gives birth to k children with probability a_k , for $k \geq 0$, independent of other individuals. Call \mathbf{a} the *offspring distribution*.

The population grows or declines from generation to generation. Let Z_n be the size (e.g., number of individuals) of the n th generation, for $n \geq 0$. Assume $Z_0 = 1$. That is, the population starts with one individual. The sequence Z_0, Z_1, \dots is a *branching process*. See Figure 4.1 for a realization of such a process through three generations.

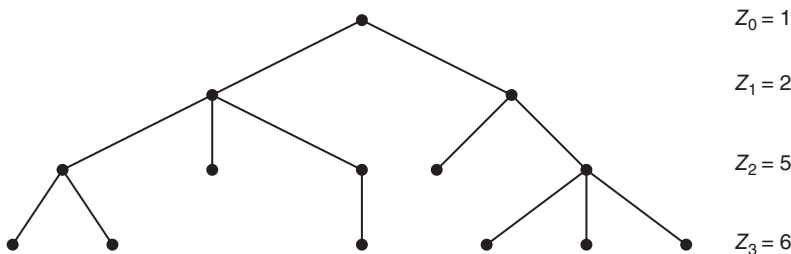


Figure 4.1 Branching process.

A branching process is a Markov chain since the size of a generation only depends on the size of the previous generation and the number of their offspring. If Z_n is given, then the size of the next generation Z_{n+1} is independent of Z_0, \dots, Z_{n-1} .

Assume $0 < a_0 < 1$. If $a_0 = 0$, then the population only grows and 0 is not in the state space. If $a_0 = 1$, then $Z_n = 0$, for all $n \geq 1$. We also assume that there is positive probability that an individual gives birth to more than one offspring, that is, $a_0 + a_1 < 1$.

Galton's first question, "What proportion of their surnames will have become extinct after r generations?" leads one to examine the recurrence and transience properties of the Markov chain.

It should be clear that 0 is an absorbing state. If a generation has no individuals, there will be no offspring. Under the initial assumptions, all other states of a branching process are transient.

Lemma 4.1. *In a branching process, all nonzero states are transient.*

Proof. If $Z_n = 0$, say the process has become *extinct* by generation n . Consider the probability that a population of size $i > 0$ goes extinct in one generation, that is, $P(Z_{n+1} = 0 | Z_n = i)$. If a generation has i individuals and the next generation has none, then each individual produced zero offspring, which occurs with probability $(a_0)^i$, by independence.

To show i is transient, we need to show that f_i , the probability of eventually hitting i for the chain started in i , is less than one. If the chain starts with i individuals, then the event that the chain eventually hits i is $\{Z_n = i \text{ for some } n \geq 1\} \subseteq \{Z_1 > 0\}$. Hence,

$$\begin{aligned} f_i &= P(Z_n = i \text{ for some } n \geq 1 | Z_0 = i) \\ &\leq P(Z_1 > 0 | Z_0 = i) \\ &= 1 - P(Z_1 = 0 | Z_0 = i) \\ &= 1 - (a_0)^i < 1, \end{aligned}$$

since $a_0 > 0$. ■

Since all nonzero states are transient and the chain has infinite state space, there are two possibilities for the long-term evolution of the process: either it gets absorbed in state 0, that is, the population eventually goes extinct, or the population grows without bound.

4.2 MEAN GENERATION SIZE

In a branching process, the size of the n th generation is the sum of the total offspring of the individuals of the previous generation. That is,

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_i, \quad (4.1)$$

where X_i denotes the number of children born to the i th person in the $(n-1)$ th generation. Because of the independence assumption, X_1, X_2, \dots is an i.i.d. sequence with common distribution \mathbf{a} . Furthermore, Z_{n-1} is independent of the X_i .

Equation (4.1) represents Z_n as a random sum of i.i.d. random variables. Results for such random sums can be applied to find the moments of Z_n .

Let $\mu = \sum_{k=0}^{\infty} k a_k$ be the mean of the offspring distribution. To find the mean of the size of the n th generation $E(Z_n)$, condition on Z_{n-1} . By the law of total expectation,

$$\begin{aligned} E(Z_n) &= \sum_{k=0}^{\infty} E(Z_n | Z_{n-1} = k) P(Z_{n-1} = k) \\ &= \sum_{k=0}^{\infty} E\left(\sum_{i=1}^{Z_{n-1}} X_i \middle| Z_{n-1} = k\right) P(Z_{n-1} = k) \end{aligned}$$

$$\begin{aligned}
&= \sum_{k=0}^{\infty} E \left(\sum_{i=1}^k X_i \middle| Z_{n-1} = k \right) P(Z_{n-1} = k) \\
&= \sum_{k=0}^{\infty} E \left(\sum_{i=1}^k X_i \right) P(Z_{n-1} = k) \\
&= \sum_{k=0}^{\infty} k \mu P(Z_{n-1} = k) = \mu E(Z_{n-1}),
\end{aligned}$$

where the fourth equality is because the X_i are independent of Z_{n-1} . Iterating the resulting recurrence relation gives

$$E(Z_n) = \mu E(Z_{n-1}) = \mu^2 E(Z_{n-2}) = \cdots = \mu^n E(Z_0) = \mu^n, \text{ for } n \geq 0,$$

since $Z_0 = 1$.

Three Cases

For the long-term expected generation size,

$$\lim_{n \rightarrow \infty} E(Z_n) = \lim_{n \rightarrow \infty} \mu^n = \begin{cases} 0, & \text{if } \mu < 1, \\ 1, & \text{if } \mu = 1, \\ \infty, & \text{if } \mu > 1. \end{cases}$$

A branching process is said to be *subcritical* if $\mu < 1$, *critical* if $\mu = 1$, and *supercritical* if $\mu > 1$. For a subcritical branching process, mean generation size declines exponentially to zero. For a supercritical process, mean generation size exhibits long-term exponential growth. The limits suggest three possible regimes depending on μ : long-term extinction, stability, and boundless growth. However, behavior of the *mean* generation size does not tell the whole story.

Insight into the evolution of a branching process is gained by simulation. We simulated 10 generations Z_0, \dots, Z_{10} of a branching process with Poisson offspring distribution, choosing three values for the Poisson mean parameter corresponding to three types of branching process: $\mu = 0.75$ (subcritical), $\mu = 1$ (critical), and $\mu = 1.5$ (supercritical). Each process was simulated five times. See the R script file **branching.R**. Results are shown in Table 4.1.

For $\mu = 0.75$, all simulated paths result in eventual extinction. Furthermore, the extinction occurs fairly rapidly.

When $\mu = 1$, all but one of the simulations in Table 4.1 become extinct by the 10th generation.

TABLE 4.1 Simulations of a Branching Process for Three Choices of μ

μ	Z_0	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
0.75	1	2	3	3	2	1	1	0	0	0	0
0.75	1	0	0	0	0	0	0	0	0	0	0
0.75	1	2	0	0	0	0	0	0	0	0	0
0.75	1	0	0	0	0	0	0	0	0	0	0
0.75	1	3	3	1	3	1	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0	0
1	1	2	0	0	0	0	0	0	0	0	0
1	3	6	6	5	6	7	8	8	8	6	5
1	1	3	4	1	2	1	0	0	0	0	0
1	1	2	1	1	2	1	2	1	0	0	0
1.5	1	2	3	10	22	41	93	173	375	763	1,597
1.5	1	1	1	1	2	4	7	9	11	19	29
1.5	1	4	5	18	34	68	127	246	521	1,011	2,065
1.5	1	1	2	0	0	0	0	0	0	0	0
1.5	1	2	5	3	2	6	9	17	18	13	19

Indeed, for a general branching process, in the subcritical and critical cases ($\mu \leq 1$), the population becomes extinct with probability 1.

In the supercritical case $\mu = 1.5$, most simulations in Table 4.1 seem to grow without bound. However, one realization goes extinct. We will see that in the general supercritical case, the probability that the population eventually dies out is less than one, but typically greater than zero.

Extinction in the Subcritical Case

Assume that Z_0, Z_1, \dots is a subcritical branching process. Let $E_n = \{Z_n = 0\}$ be the event that the population is extinct by generation n , for $n \geq 1$. Let E be the event that the population is ultimately extinct. Then,

$$E = \{Z_n = 0, \text{ for some } n \geq 1\} = \bigcup_{n=1}^\infty E_n,$$

and $E_1 \subseteq E_2 \subseteq \dots$. It follows that the probability that the population eventually goes extinct is

$$P(E) = P\left(\bigcup_{n=1}^\infty E_n\right) = \lim_{n \rightarrow \infty} P(E_n) = \lim_{n \rightarrow \infty} P(Z_n = 0). \tag{4.2}$$

The probability that the population is extinct by generation n is

$$\begin{aligned}
 P(Z_n = 0) &= 1 - P(Z_n \geq 1) \\
 &= 1 - \sum_{k=1}^{\infty} P(Z_n = k) \\
 &\geq 1 - \sum_{k=1}^{\infty} kP(Z_n = k) \\
 &= 1 - E(Z_n) = 1 - \mu^n.
 \end{aligned}$$

Taking limits gives

$$P(E) = \lim_{n \rightarrow \infty} P(Z_n = 0) \geq \lim_{n \rightarrow \infty} 1 - \mu^n = 1,$$

since $\mu < 1$. Thus, $P(E) = 1$. With probability 1, a subcritical branching process eventually goes extinct.

■ **Example 4.1** Subcritical branching processes have been used to model the spread of infections and disease in highly vaccinated populations. Farrington and Grant (1999) cite several examples, including the spread of measles and mumps, the outbreak of typhoidal salmonellae reported in Scotland in 1967–1990, and outbreaks of human monkeypox virus in past decades. Becker (1974) finds evidence of subcriticality in European smallpox data from 1950 to 1970.

Often the goal of these studies is to use data on observed outbreaks to estimate the unknown mean offspring parameter μ as well as the number of generations of spread until extinction. ■

Variance of Generation Size

To explore the process of extinction in the critical and supercritical cases ($\mu \geq 1$), we first consider the variance of the size of the n th generation $\text{Var}(Z_n)$. Let σ^2 denote the variance of the offspring distribution. By the law of total variance,

$$\text{Var}(Z_n) = \text{Var}(E(Z_n|Z_{n-1})) + E(\text{Var}(Z_n|Z_{n-1})).$$

We have shown that

$$E(Z_n|Z_{n-1} = k) = E\left(\sum_{i=1}^k X_i\right) = \sum_{i=1}^k E(X_i) = \mu k,$$

which gives $E(Z_n|Z_{n-1}) = \mu Z_{n-1}$. Similarly, $\text{Var}(Z_n|Z_{n-1}) = \sigma^2 Z_{n-1}$, since

$$\text{Var}(Z_n|Z_{n-1} = k) = \text{Var}\left(\sum_{i=1}^k X_i\right) = \sum_{i=1}^k \text{Var}(X_i) = \sigma^2 k,$$

using the independence of the X_i . Applying the law of total variance,

$$\begin{aligned} \text{Var}(Z_n) &= \text{Var}(\mu Z_{n-1}) + E(\sigma^2 Z_{n-1}) \\ &= \mu^2 \text{Var}(Z_{n-1}) + \sigma^2 \mu^{n-1}, \text{ for } n \geq 1. \end{aligned} \quad (4.3)$$

With $\text{Var}(Z_0) = 0$, Equation (4.3) yields

$$\begin{aligned} \text{Var}(Z_1) &= \mu^2 \text{Var}(Z_0) + \sigma^2 = \sigma^2, \\ \text{Var}(Z_2) &= \mu^2 \text{Var}(Z_1) + \sigma^2 \mu = \sigma^2 \mu(1 + \mu), \text{ and} \\ \text{Var}(Z_3) &= \mu^2 \text{Var}(Z_2) + \sigma^2 \mu^2 = \sigma^2 \mu^2(1 + \mu + \mu^2). \end{aligned}$$

The general pattern, proved by induction on n , gives

$$\text{Var}(Z_n) = \sigma^2 \mu^{n-1} \sum_{k=0}^{n-1} \mu^k = \begin{cases} n\sigma^2, & \text{if } \mu = 1, \\ \sigma^2 \mu^{n-1}(\mu^n - 1)/(\mu - 1), & \text{if } \mu \neq 1. \end{cases}$$

In the subcritical case, both the mean and variance of generation size tend to 0.

In the critical case, the mean size of every generation is one, but the variance is a linearly growing function of n .

In the supercritical case, the variance grows exponentially large. The potentially large difference between the mean μ^n and variance suggests that in some cases both extinction and boundless growth are possible outcomes.

To explore the issue more carefully, we will find the probability of ultimate extinction when $\mu \geq 1$. First, however, new tools are needed.

4.3 PROBABILITY GENERATING FUNCTIONS

For a discrete random variable X taking values in $\{0, 1, \dots\}$, the *probability generating function of X* is the function

$$\begin{aligned} G(s) &= E(s^X) = \sum_{k=0}^{\infty} s^k P(X = k) \\ &= P(X = 0) + sP(X = 1) + s^2P(X = 2) + \dots \end{aligned}$$

The function is a power series whose coefficients are probabilities. Observe that $G(1) = 1$. The series converges absolutely for $|s| \leq 1$. To emphasize the underlying random variable X , we may write $G(s) = G_X(s)$.

The generating function represents the distribution of a discrete random variable as a power series. If two power series are equal, then they have the same coefficients. Hence, if two discrete random variables X and Y have the same probability generating function, that is, $G_X(s) = G_Y(s)$ for all s , then X and Y have the same distribution.

■ **Example 4.2** Let X be uniformly distributed on $\{0, 1, 2\}$. Find the probability generating function of X .

Solution

$$G(s) = E(s^X) = \frac{1}{3} + s\left(\frac{1}{3}\right) + s^2\left(\frac{1}{3}\right) = \frac{1}{3}(1 + s + s^2).$$

■

■ **Example 4.3** Assume that X has a geometric distribution with parameter p . Find the probability generating function of X .

Solution

$$G(s) = E(s^X) = \sum_{k=1}^{\infty} s^k p(1-p)^{k-1} = sp \sum_{k=1}^{\infty} (s(1-p))^{k-1} = \frac{sp}{1-s(1-p)},$$

for $|s| < 1$.

■

Probabilities for X can be obtained from the generating function by successive differentiation. We have that

$$G(0) = P(X = 0),$$

$$G'(0) = \sum_{k=1}^{\infty} k s^{k-1} P(X = k) \Big|_{s=0} = P(X = 1),$$

$$G''(0) = \sum_{k=2}^{\infty} k(k-1) s^{k-2} P(X = k) \Big|_{s=0} = 2P(X = 2),$$

and so on. In general,

$$G^{(j)}(0) = \sum_{k=j}^{\infty} k(k-1) \cdots (k-j+1) s^{k-j} P(X = k) \Big|_{s=0} = j! P(X = j),$$

and thus

$$P(X = j) = \frac{G^{(j)}(0)}{j!}, \quad \text{for } j = 0, 1, \dots,$$

where $G^{(j)}$ denotes the j th derivative of G .

■ **Example 4.4** A random variable X has probability generating function

$$G(s) = (1 - p + sp)^n.$$

Find the distribution of X .

Solution We have $P(X = 0) = G(0) = (1 - p)^n$. For $1 \leq j \leq n$, the j th derivative of G is

$$G^{(j)}(s) = n(n-1) \cdots (n-j+1)p^j(1-p+sp)^{n-j},$$

which gives

$$\begin{aligned} P(X = j) &= \frac{G^{(j)}(0)}{j!} \\ &= \frac{n(n-1) \cdots (n-j+1)}{j!} p^j (1-p)^{n-j} \\ &= \binom{n}{j} p^j (1-p)^{n-j}. \end{aligned}$$

For $j > n$, $G^{(j)}(0) = 0$, and thus $P(X = j) = 0$. We see that X has a binomial distribution with parameters n and p . ■

Sums of Independent Random Variables

Generating functions are useful tools for working with sums of independent random variables. Assume that X_1, \dots, X_n are independent. Let $Z = X_1 + \cdots + X_n$. The probability generating function of Z is

$$\begin{aligned} G_Z(s) &= E(s^Z) = E(s^{X_1 + \cdots + X_n}) \\ &= E\left(\prod_{k=1}^n s^{X_k}\right) = \prod_{k=1}^n E(s^{X_k}) \\ &= G_{X_1}(s) \cdots G_{X_n}(s), \end{aligned}$$

where the fourth equality is by independence. The generating function of an independent sum is the product of the individual generating functions. If the X_i are also identically distributed, then

$$G_Z(s) = G_{X_1}(s) \cdots G_{X_n}(s) = [G_X(s)]^n,$$

where X is a random variable with the same distribution as the X_i .

■ **Example 4.5** In Example 4.4, it is shown that the generating function of a binomial random variable with parameters n and p is $G(s) = (1 - p + ps)^n$. Here is a derivation using the fact that a sum of i.i.d. Bernoulli random variables has a binomial distribution.

Solution Let X_1, \dots, X_n be an i.i.d. sequence of Bernoulli random variables with parameter p . The common generating function of the X_i is

$$G(s) = E(s^{X_i}) = s^0 P(X_i = 0) + s^1 P(X_i = 1) = (1 - p) + sp.$$

The sum $Z = X_1 + \dots + X_n$ has a binomial distribution with parameters n and p . The probability generating function of Z is thus

$$G_Z(s) = [G(s)]^n = (1 - p + ps)^n. \quad \blacksquare$$

Moments

The probability generating function of X can be used to find the mean, variance, and higher moments of X . Observe that

$$G'(1) = E(Xs^{X-1}) \Big|_{s=1} = E(X).$$

Also,

$$G''(1) = E(X(X-1)s^{X-2}) \Big|_{s=1} = E(X(X-1)) = E(X^2) - E(X),$$

which gives

$$\begin{aligned} \text{Var}(X) &= E(X^2) - E(X)^2 = (E(X^2) - E(X)) + E(X) - E(X)^2 \\ &= G''(1) + G'(1) - G'(1)^2. \end{aligned}$$

Example 4.6 For a geometric random variable with parameter p , the generating function is

$$G(s) = \frac{sp}{1 - s(1 - p)},$$

as shown in Example 4.3. Use the generating function to find the mean and variance of the geometric distribution.

Solution For the mean,

$$G'(s) = \frac{p}{(1 - s(1 - p))^2},$$

which gives $E(X) = G'(1) = 1/p$. For the variance,

$$G''(s) = \frac{2p(1 - p)}{(1 - s(1 - p))^3}, \quad \text{and} \quad G''(1) = \frac{2(1 - p)p}{p^2}.$$

This gives

$$\text{Var}(X) = G''(1) + G'(1) - G'(1)^2 = \frac{2(1 - p)p}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1 - p}{p^2}.$$

We summarize some key properties of probability generating functions. ■

Properties of Probability Generating Function

1. Let $G(s) = E(s^X)$ be the probability generating function of a discrete random variable X . Then,
 - (a) $G(1) = 1$,
 - (b) $P(X = k) = G^{(k)}(0)/k!$, for $k \geq 0$,
 - (c) $E(X) = G'(1)$,
 - (d) $\text{Var}(X) = G''(1) + G'(1) - G'(1)^2$.
2. If X and Y are random variables such that $G_X(s) = G_Y(s)$ for all s , then X and Y have the same distribution.
3. If X and Y are independent, then $G_{X+Y}(s) = G_X(s)G_Y(s)$.

4.4 EXTINCTION IS FOREVER

Probability generating functions are especially useful for analyzing branching processes. We use them to find the probability that a branching process eventually goes extinct.

For $n \geq 0$, let

$$G_n(s) = \sum_{k=0}^{\infty} s^k P(Z_n = k)$$

be the generating function of the n th generation size Z_n . Let

$$G(s) = \sum_{k=0}^{\infty} s^k a_k$$

be the generating function of the offspring distribution. We have

$$G_n(s) = E(s^{Z_n}) = E\left(s^{\sum_{k=1}^{Z_{n-1}} X_k}\right) = E\left(E\left(s^{\sum_{k=1}^{Z_{n-1}} X_k} \mid Z_{n-1}\right)\right),$$

where the last equality is by the law of total expectation. From the independence of Z_{n-1} and the X_k ,

$$\begin{aligned} E\left(s^{\sum_{k=1}^{Z_{n-1}} X_k} \mid Z_{n-1} = z\right) &= E\left(s^{\sum_{k=1}^z X_k} \mid Z_{n-1} = z\right) \\ &= E\left(s^{\sum_{k=1}^z X_k}\right) = E\left(\prod_{k=1}^z s^{X_k}\right) \\ &= \prod_{k=1}^z E(s^{X_k}) = [G(s)]^z, \end{aligned}$$

for all z . This gives

$$E \left(s^{\sum_{k=1}^{Z_{n-1}} X_k} | Z_{n-1} \right) = [G(s)]^{Z_{n-1}}.$$

Taking expectations,

$$G_n(s) = E \left(G(s)^{Z_{n-1}} \right) = G_{n-1}(G(s)), \quad \text{for } n \geq 1.$$

The probability generating function of Z_n is the composition of the generating function of Z_{n-1} and the generating function of the offspring distribution.

Observe that $G_0(s) = s$, and $G_1(s) = G_0(G(s)) = G(s)$. From the latter we see that the distribution of Z_1 is the offspring distribution \mathbf{a} .

Continuing,

$$G_2(s) = G_1(G(s)) = G(G(s)) = G(G_1(s)),$$

and

$$G_3(s) = G_2(G(s)) = G(G(G(s))) = G(G_2(s)).$$

In general,

$$G_n(s) = G_{n-1}(G(s)) = \underbrace{G(\cdots G(G(s)) \cdots)}_{n\text{-fold}} = G(G_{n-1}(s)). \quad (4.4)$$

The generating function of Z_n is the n -fold composition of the offspring distribution generating function.

Equation (4.4) is typically not useful for computing the actual distribution of Z_n . For an arbitrary offspring distribution, the distribution of Z_n will be complicated with no tractable closed-form expression. However, the equation is central to the proof of the following theorem, which characterizes the extinction probability for a branching process.

Extinction Probability

Theorem 4.2. *Given a branching process, let G be the probability generating function of the offspring distribution. Then, the probability of eventual extinction is the smallest positive root of the equation $s = G(s)$.*

If $\mu \leq 1$, that is, in the subcritical and critical cases, the extinction probability is equal to 1.

Remark: We have already shown that in the subcritical $\mu < 1$ case, the population goes extinct with probability 1. The theorem gives that this is also true for $\mu = 1$, even though for each generation the expected generation size is $E(Z_n) = \mu^n = 1$. For the supercritical case $\mu > 1$, the expected generation size Z_n grows without bound.

However, the theorem gives that even in this case there is positive probability of eventual extinction.

Before proving Theorem 4.2, we offer some examples of its use. Let e denote the probability of eventual extinction.

■ **Example 4.7** Find the extinction probability for a branching process with offspring distribution $\mathbf{a} = (1/6, 1/2, 1/3)$.

Solution The mean of the offspring distribution is

$$\mu = 0(1/6) + 1(1/2) + 2(1/3) = 7/6 > 1,$$

so this is the supercritical case. The offspring generating function is

$$G(s) = \frac{1}{6} + \frac{s}{2} + \frac{s^2}{3}.$$

Solving

$$s = G(s) = \frac{1}{6} + \frac{s}{2} + \frac{s^2}{3}$$

gives the quadratic equation $s^2/3 - s/2 + 1/6 = 0$, with roots $s = 1$ and $s = 1/2$. The smallest positive root is the probability of eventual extinction $e = 1/2$. ■

■ **Example 4.8** A branching process has offspring distribution

$$a_k = (1 - p)^k p, \quad \text{for } k = 0, 1, \dots$$

Find the extinction probability in the supercritical case.

Solution The offspring distribution is a variant of the geometric distribution. The generating function is

$$G(s) = \sum_{k=0}^{\infty} s^k (1 - p)^k p = p \sum_{k=0}^{\infty} (s(1 - p))^k = \frac{p}{1 - s(1 - p)}, \quad \text{for } |s(1 - p)| < 1.$$

The mean of the offspring distribution is

$$\mu = G'(1) = \left. \frac{p(1 - p)}{(1 - s(1 - p))^2} \right|_{s=1} = \frac{1 - p}{p}.$$

The supercritical case $\mu > 1$ corresponds to $p < 1/2$.

To find the extinction probability, solve

$$s = G(s) = \frac{p}{1 - s(1 - p)},$$

which gives the quadratic equation

$$(1-p)s^2 - s + p = 0,$$

with roots

$$s = \frac{1 \pm \sqrt{1 - (4(1-p)p)}}{2(1-p)} = \frac{1 \pm (1-2p)}{2(1-p)}.$$

The roots are 1 and $p/(1-p)$. For $0 < p < 1/2$, the smaller root is $p/(1-p) = 1/\mu$, that is, $e = 1/\mu$. ■

We explore the extinction probability result in Example 4.8 with the use of simulation. Let $p = 1/4$. We simulated a branching process with offspring distribution $a_k = (3/4)^k(1/4)$, for $k \geq 0$. The process is supercritical with $\mu = 3$. Results are collected in Table 4.2. Four of the 12 runs went extinct by time $n = 10$. The exact extinction probability is $e = 1/3$. (The fact that 4 out of 12 is exactly $1/3$ is, we assure the reader, pure coincidence.)

R: Simulating the Extinction Probability

The branching process is simulated 10,000 times, keeping track of the number of times the process goes extinct by the 10th generation. See the file **branching.R**.

```
> branch(10, 1/4)
 1   4   6  16  71  205  569  1559 4588 13726 40800

> trials <- 10000
> simlist <- replicate(trials, branch(10, 1/4)[11])
# Estimate of extinction probability
> sum(simlist==0)/trials
[1] 0.332
```

The function `branch(n, p)` simulates n steps of a branching process whose offspring distribution is geometric with parameter p . The `replicate` command repeats the simulation 10,000 times, storing the outcome of Z_{10} for each trial in the vector `simlist`. The proportion of 0s in `simlist` estimates the extinction probability e .

Our conclusions are slightly biased since we assume that if extinction takes place it will occur by time $n = 10$. Of course, extinction could occur later. However, it appears from the simulations that if extinction occurs it happens very rapidly.

TABLE 4.2 Simulation of a Supercritical Branching Process, with $\mu = 3$. Four of the 12 runs go extinct by the 10th generation.

Z_0	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
1	4	25	97	394	1160	3475	10685	31885	95757	287130
1	1	1	3	7	11	47	165	515	1525	4689
1	10	37	115	350	1124	3455	10073	29896	88863	267386
1	1	1	2	2	3	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
1	8	31	71	248	779	2282	6864	19895	59196	178171
1	0	0	0	0	0	0	0	0	0	0
1	4	7	13	34	106	380	1123	3385	10200	30090
1	16	49	163	447	1284	3794	11592	34626	104390	312704
1	1	1	5	16	51	155	559	1730	5378	15647
1	1	3	31	79	267	883	2637	8043	23970	71841
1	0	0	0	0	0	0	0	0	0	0

■ **Example 4.9 (Lotka’s estimate of the extinction probability)** One of the earliest applications of branching processes is contained in the work of Alfred Lotka, considered the father of demographic analysis, who estimated the probability that a male line of descent would ultimately become extinct. Based on the 1920 census data, Lotka fitted the distribution of male offspring to a *zero-adjusted* geometric distribution of the form

$$a_0 = 0.48235 \quad \text{and} \quad a_k = (0.2126)(0.5893)^{k-1}, \quad \text{for } k \geq 1.$$

The generating function of the offspring distribution is

$$G(s) = 0.48235 + 0.2126 \sum_{k=1}^{\infty} (0.5893)^{k-1} s^k = 0.48235 + \frac{(0.2126)s}{1 - (0.5893)s}.$$

Lotka found the extinction probability as the numerical solution to $G(s) = s$, giving the value $e = 0.819$.

The mean of the male offspring distribution is $\mu = 1.26$. It is interesting that despite a mean number of children (sons and daughters) per individual of about 2.5, the probability of extinction of family surnames is over 80% See Lotka (1931) and Hull (2001). ■

■ **Example 4.10** A *worm* is a self-replicating computer virus, which exploits computer network security vulnerabilities to spread itself. The *Love Letter* was a famous worm, which attacked tens of millions of Windows computers in 2000. It was spread as an attachment to an email message with the subject line ILOVEYOU. When users clicked on the attachment the worm automatically downloaded onto their machines.

Sellke et al. (2008) modeled the spread of computer worms as a branching process. The worms they analyzed are spread by randomly scanning from the 2^{32} current IP addresses to find a vulnerable host. Let V denote the total number of vulnerable hosts. Then, $p = V/2^{32}$ is the probability of finding a vulnerable host in one scan. If a worm scans at most M hosts, then infected hosts represent the individuals of a branching process whose offspring distribution is binomial with parameters M and p . Since the mean of the offspring distribution is Mp , it follows that the spread of the worm will eventually die out, with probability 1, if $Mp \leq 1$, or $M \leq 1/p$.

Typically, M is large and p is small and thus the offspring binomial distribution is well approximated by a Poisson distribution with parameter $\lambda = Mp$. The total number of infected hosts T before the worm eventually dies out is the *total progeny* of the branching process. See Exercises 4.24 and 4.25, where the mean and variance of the total progeny of a branching process are derived. If the worm starts out with I infected hosts then the mean and variance of the total number of infected hosts before the virus dies out is

$$E(T) = \frac{I}{1 - \lambda} \quad \text{and} \quad \text{Var}(T) = \frac{I}{(1 - \lambda)^3}.$$

■

Proof of Extinction Probability Theorem 4.2

Proof. Let $e_n = P(Z_n = 0)$ denote the probability that the population goes extinct by generation n . We have

$$\begin{aligned} e_n &= P(Z_n = 0) = G_n(0) = G(G_{n-1}(0)) \\ &= G(P(Z_{n-1} = 0)) = G(e_{n-1}), \end{aligned} \tag{4.5}$$

for $n \geq 1$. From Equation (4.2), $e_n \rightarrow e$, as $n \rightarrow \infty$. Taking limits on both sides of Equation (4.5), as $n \rightarrow \infty$, and using the fact that the probability generating function is continuous, gives $e = G(e)$. Thus, e is a root of the equation $s = G(s)$.

Let x be a positive solution of $s = G(s)$. We need to show that $e \leq x$. Since $G(s) = \sum_k s^k P(X = k)$ is an increasing function on $(0, 1]$, and $0 < x$,

$$e_1 = P(Z_1 = 0) = G_1(0) = G(0) \leq G(x) = x.$$

By induction, assuming $e_k \leq x$, for all $k < n$,

$$e_n = P(Z_n = 0) = G_n(0) = G(G_{n-1}(0)) = G(e_{n-1}) \leq G(x) = x.$$

Taking limits as $n \rightarrow \infty$, gives $e \leq x$. This proves the first part of the theorem.

The remainder of the theorem is essentially revealed by Figure 4.2. Consider the intersection of the graph of $y = G(s)$ with the line $y = s$ on the interval $[0, 1]$. Observe

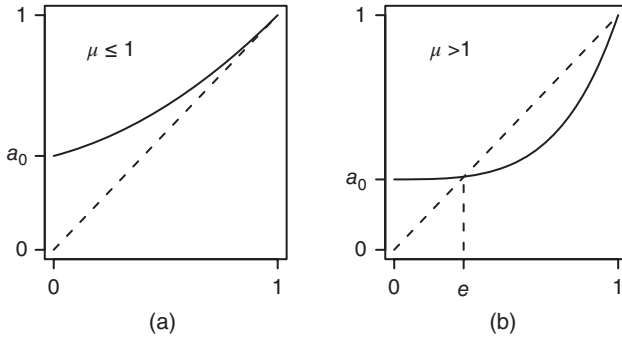


Figure 4.2 Graph of $G(s)$.

that $G(0) = a_0$ and $G(1) = 1$. Furthermore, the continuous and differentiable function G is convex (concave up) as

$$G''(s) = \sum_{k=2}^{\infty} k(k-1)s^{k-2}P(X = k) > 0.$$

It follows that the graph of $y = G(s)$ can intersect the line $y = s$ at either one or two points.

What distinguishes the two cases is the derivative of $G(s)$ at $s = 1$. Recall that $G'(1) = \mu$.

- (i) If $\mu = G'(1) \leq 1$, we are in the setting of Figure 4.2(b). Since

$$G'(s) = \sum_{k=1}^{\infty} ks^{k-1}P(X = k) > 0$$

is a strictly increasing function of s , we have $G'(s) < G'(1) = 1$, for $0 < s < 1$. Let $h(s) = s - G(s)$. Then, $h'(s) = 1 - G'(s) > 0$, for $0 < s < 1$. Since h is increasing and $h(1) = 0$, it follows that $h(s) < 0$, for $0 < s < 1$. That is, $s < G(s)$. Hence, the graph of $G(s)$ lies above the line $y = s$, for $0 < s < 1$, and $s = 1$ is the only point of intersection. Hence, the extinction probability is $e = 1$.

- (ii) If $\mu = G'(1) > 1$, we are in the setting of Figure 4.2(b). Here,

$$h(0) = 0 - G(0) = -a_0 < 0.$$

Also, $h'(1) = 1 - G'(1) = 1 - \mu < 0$, thus $h(s)$ is decreasing at $s = 1$. Since $h(1) = 0$, there is some $0 < t < 1$ such that $h(t) > 0$. It follows that there is a number e between 0 and 1 such that $h(e) = 0$. That is, $e = G(e)$. This is the desired extinction probability. ■

EXERCISES

- 4.1** Consider a branching process with offspring distribution $\mathbf{a} = (a, b, c)$, where $a + b + c = 1$. Let \mathbf{P} be the Markov transition matrix. Exhibit the first three rows of \mathbf{P} . That is, find P_{ij} for $i = 0, 1, 2$ and $j = 0, 1, \dots$
- 4.2** Find the probability generating function of a Poisson random variable with parameter λ . Use the pgf to find the mean and variance of the Poisson distribution.
- 4.3** Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$. Assume that X and Y are independent. Use probability generating functions to find the distribution of $X + Y$.
- 4.4** If X is a negative binomial distribution with parameters r and p , then X can be written as the sum of r i.i.d. geometric random variables with parameter p . Use this fact to find the pgf of X . Then, use the pgf to find the mean and variance of the negative binomial distribution.
- 4.5** The k th factorial moment of a random variable X is

$$E(X(X-1)\cdots(X-k+1)) = E\left(\frac{X!}{(X-k)!}\right), \quad \text{for } k \geq 0.$$

- (a) Given the probability generating function G of X , show how to find the k th factorial moment of X .
- (b) Find the k th factorial moment of a binomial random variable with parameters n and p .
- 4.6** Let X_1, X_2, \dots be a sequence of i.i.d. Bernoulli random variables with parameter p . Let N be a Poisson random variable with parameter λ , which is independent of the X_i .
- (a) Find the probability generating function of $Z = \sum_{i=1}^N X_i$.
- (b) Use (a) to identify the probability distribution of Z .
- 4.7** Give the probability generating function for an offspring distribution in which an individual either dies, with probability $1 - p$, or gives birth to three children, with probability p . Also find the mean and variance of the number of children in the fourth generation.
- 4.8** If X is a discrete random variable with generating function G . Show that

$$P(X \text{ is even}) = \frac{1 + G(-1)}{2}.$$

- 4.9** Let Z_0, Z_1, \dots be a branching process whose offspring distribution mean is μ . Let $Y_n = Z_n/\mu^n$, for $n \neq 0$. Show that $E(Y_{n+1}|Y_n) = Y_n$.

- 4.10** Show by induction that for $\mu \neq 1$,

$$\text{Var}(Z_n) = \sigma^2 \mu^{n-1} \frac{\mu^n - 1}{\mu - 1}.$$

- 4.11** Use the generating function representation of Z_n in Equation (4.4) to find $E(Z_n)$.
- 4.12** A branching process has offspring distribution $\mathbf{a} = (1/4, 1/4, 1/2)$. Find the following:
- μ .
 - $G(s)$.
 - The extinction probability.
 - $G_2(s)$.
 - $P(Z_2 = 0)$.
- 4.13** Use numerical methods to find the extinction probability for a branching process with Poisson offspring distribution with parameter $\lambda = 1.5$.
- 4.14** A branching process has offspring distribution with $a_0 = p$, $a_1 = 1 - p - q$, and $a_2 = q$. For what values of p and q is the process supercritical? In the supercritical case, find the extinction probability.
- 4.15** Assume that the offspring distribution is uniform on $\{0, 1, 2, 3, 4\}$. Find the extinction probability.
- 4.16** Consider a branching process where $Z_0 = k$. That is, the process starts with k individuals. Let $G(s)$ be the probability generating function of the offspring distribution. Let $G_n(s)$ be the probability generating function of Z_n for $n = 0, 1, \dots$
- Find the probability generating function $G_1(s)$ in terms of $G(s)$.
 - True or False: $G_{n+1}(s) = G_n(G(s))$, for $n = 1, 2, \dots$
 - True or False: $G_{n+1}(s) = G(G_n(s))$, for $n = 1, 2, \dots$
- 4.17** For $0 < p < 1$, let $\mathbf{a} = (1 - p, 0, p)$ be the offspring distribution of a branching process. Each individual in the population can have either two or no offspring. Assume that the process starts with two individuals.
- Find the extinction probability.
 - Write down the general term P_{ij} for the Markov transition matrix of the branching process.
- 4.18** Consider a branching process with offspring distribution
- $$\mathbf{a} = (p^2, 2p(1 - p), (1 - p)^2), \quad \text{for } 0 < p < 1.$$
- The offspring distribution is binomial with parameters 2 and $1 - p$. Find the extinction probability.
- 4.19** Let $T = \min\{n : Z_n = 0\}$ be the time of extinction for a branching process. Show that $P(T = n) = G_n(0) - G_{n-1}(0)$, for $n \geq 1$.
- 4.20** Consider the offspring distribution defined by $a_k = (1/2)^{k+1}$, for $k \geq 0$.
- Find the extinction probability.
 - Show by induction that

$$G_n(s) = \frac{n - (n - 1)s}{n + 1 - ns}.$$

- See Exercise 4.19. Find the distribution of the time of extinction.

- 4.21** The *linear fractional case* is one of the few branching process examples in which the generating function $G_n(s)$ can be explicitly computed. For $0 < p < 1$, let

$$a_0 = \frac{1 - c - p}{1 - p}, \quad a_k = cp^{k-1}, \quad \text{for } k = 1, 2, \dots,$$

where $0 < c < 1 - p$ is a parameter. The offspring distribution is a geometric distribution rescaled at 0.

- (a) Find μ , the mean of the offspring distribution.
 (b) Assume that $\mu = 1$. Show, by induction, that

$$G_n(s) = \frac{np - (np + p - 1)s}{1 - p + np - nps}.$$

- (c) For $\mu > 1$, Athreya and Ney (1972) show

$$G_n(s) = \frac{(\mu^n e - 1)s + e(1 - \mu^n)}{(\mu^n - 1)s + e - \mu^n},$$

where $e = (1 - c - p)/(p(1 - p))$ is the extinction probability. See Example 4.9. Observe that Lotka's model falls in the linear fractional case. For Lotka's data, find the probability that a male line of descent goes extinct by the third generation.

- 4.22** *Linear fractional case, continued.* A rumor-spreading process evolves as follows. At time 0, one person has heard a rumor. At each discrete unit of time every person who has heard the rumor decides how many people to tell according to the following mechanism. Each person flips a fair coin. If heads, they tell no one. If tails, they proceed to roll a fair die until 5 appears. The number of rolls needed determines how many people they will tell the rumor.
- (a) After four *generations*, how many people, on average, have heard the rumor?
 (b) Find the probability that the rumor-spreading process will stop after four generations.
 (c) Find the probability that the rumor-spreading process will eventually stop.
- 4.23** Let \mathbf{a} be an offspring distribution with generating function G . Let X be a random variable with distribution \mathbf{a} . Let Z be a random variable whose distribution is that of X conditional on $X > 0$. That is, $P(Z = k) = P(X = k | X > 0)$. Find the generating function of Z in terms of G .
- 4.24** Let $T_n = Z_0 + Z_1 + \dots + Z_n$ be the total number of individuals up through generation n . Let $T = \lim_{n \rightarrow \infty} T_n$ be the *total progeny* of the branching process. Find $E(T)$ for the subcritical, critical, and supercritical cases.
- 4.25** *Total progeny, continued.* Let $\phi_n(s) = E(s^{T_n})$ be the probability generating function of T_n , as defined in Exercise 4.24.

- (a) Show that ϕ_n satisfies the recurrence relation

$$\phi_n(s) = sG(\phi_{n-1}(s)), \quad \text{for } n = 1, 2, \dots,$$

where $G(s)$ is the pgf of the offspring distribution. Hint: Condition on Z_1 and use Exercise 4.16(a).

- (b) From (a), argue that

$$\phi(s) = sG(\phi(s)),$$

where $\phi(s)$ is the pgf of the total progeny T .

- (c) Use (b) to find the mean of T in the subcritical case.

4.26 In a lottery game, three winning numbers are chosen uniformly at random from $\{1, \dots, 100\}$, sampling without replacement. Lottery tickets cost \$1 and allow a player to pick three numbers. If a player matches the three winning numbers they win the jackpot prize of \$1,000. For matching exactly two numbers, they win \$15. For matching exactly one number they win \$3.

- (a) Find the distribution of net winnings for a random lottery ticket. Show that the expected value of the game is -70.8 cents.
- (b) *Parlaying* bets in a lottery game occurs when the winnings on a lottery ticket are used to buy tickets for future games. Hoppe (2007) analyzes the effect of parlaying bets on several lottery games. Assume that if a player matches either one or two numbers they parlay their bets, buying respectively 3 or 15 tickets for the next game. The number of tickets obtained by parlaying can be considered a branching process. Find the mean of the offspring distribution and show that the process is subcritical.
- (c) See Exercise 4.19. Let T denote the duration of the process, that is, the length of the parlay. Find $P(T = k)$, for $k = 1, \dots, 4$.
- (d) Hoppe shows that the probability that a single parlayed ticket will ultimately win the jackpot is approximately $p/(1 - m)$, where p is the probability that a single ticket wins the jackpot, and m is the mean of the offspring distribution of the associated branching process. Find this probability and show that the parlaying strategy increases the probability that a ticket will ultimately win the jackpot by slightly over 40%.

4.27 Consider a branching process whose offspring distribution is Bernoulli with parameter p .

- (a) Find the probability generating function for the n th generation size Z_n . Describe the distribution of Z_n .
- (b) For $p = 0.9$, find the extinction probability and the expectation of total progeny.

4.28 In a *branching process with immigration*, a random number of immigrants W_n is independently added to the population at the n th generation.

- (a) Let H_n be the probability generating function of W_n . If G_n is the generating function of the size of the n th generation, show that

$$G_n(s) = G_{n-1}(G(s))H_n(s).$$

- (b) Assume that the offspring distribution is Bernoulli with parameter p , and the immigration distribution is Poisson with parameter λ . Find the generating function $G_n(s)$, and show that

$$\lim_{n \rightarrow \infty} G_n(s) = e^{-\lambda(1-s)/(1-p)}.$$

What can you conclude about the limiting distribution of generation size?

- 4.29** R: Examine the proof of Theorem 4.2 and observe that

$$e_n = G(e_{n-1}), \quad \text{for } n \geq 1, \quad (4.6)$$

where $e_n = P(Z_n = 0)$ is the probability that the population goes extinct by generation n . Since $e_n \rightarrow e$, as $n \rightarrow \infty$, Equation (4.6) is the basis for a numerical, recursive method to approximate the extinction probability in the supercritical case. To find e :

1. Initialize with $e_0 \in (0, 1)$.
2. Successively compute $e_n = G(e_{n-1})$, for $n \geq 1$.
3. Set $e = e_n$, for large n .

Convergence can be shown to be exponentially fast, so that n can often be taken to be relatively small (e.g., $n \approx 10$ -20). Use this numerical method to find the extinction probability for the following cases.

- (a) $a_0 = 0.8, a_4 = 0.1, a_9 = 0.1$.
- (b) Offspring distribution is uniform on $\{0, 1, \dots, 10\}$.
- (c) $a_0 = 0.6, a_3 = 0.2, a_6 = 0.1, a_{12} = 0.1$.

- 4.30** R: Simulate the branching process in Exercise 4.12. Use your simulation to estimate the extinction probability e .
- 4.31** R: Simulating a branching process whose offspring distribution is uniformly distributed on $\{0, 1, 2, 3, 4\}$.
- (a) Use your simulation to estimate the probability that the process goes extinct by the third generation. Compare with the exact result obtained by numerical methods.
 - (b) See Exercise 4.15. Use your simulation to estimate the extinction probability e . Assume that if the process goes extinct it will do so by the 10th generation with high probability.
- 4.32** R: Simulate the branching process with immigration in Exercise 4.28(b), with $p = 3/4$ and $\lambda = 1.2$. Illustrate the limit result in Exercise 4.28(c) with $n = 100$.

- 4.33** R: Simulate the total progeny for a branching process whose offspring distribution is Poisson with parameter $\lambda = 0.60$. Estimate the mean and variance of the total progeny distribution.
- 4.34** R: Based on the numerical algorithm in Exercise 4.29, write an R function `extinct(offspring)` to find the extinction probability for any branching process with a finite offspring distribution.