

## BRANCHING PROCESSES

### 6.1. Probability generating functions

Branching processes are Markov chains, but because of their unusual structure they are best handled using probability generating functions. It has been a long time since Hong talked about generating functions. So, let me “remind” you of some important properties of probability generating functions that I will need.

**Definition 6.1:** If  $X$  is an integer-valued random variable with probability distribution

$$\Pr \{ X = k \} = p_k, \quad k = 0, 1, 2, \dots \quad (6.1)$$

then the power series

$$\begin{aligned} F(s) &= E(s^X) = \sum_{k=0}^{\infty} p_k s^k \\ &= p_0 + p_1 s + p_2 s^2 + \dots \end{aligned} \quad (6.2)$$

is the *probability generating function* of  $X$ .

Probability generating functions are quite old. Their use goes back to de Moivre, in 1738, and Laplace, in 1782. Probability generating functions are also closely related to what engineers call  $z$ -transforms.

Since every  $p_k$  is less than or equal to one, the probability generating function converges for all  $|s| < 1$  (by comparison with the geometric series). Since the  $p_n$  sum to one,  $F(s)$  also converges for  $s = 1$ . This is enough to guarantee that our power series converges uniformly for  $|s| \leq 1$  and that our generating function is a continuous function on this interval. There are a number of other properties of the probability generating function  $F(s)$  that are worth highlighting:

- (a) The probability that  $X = 0$ ,  $p_0$ , is given by

$$p_0 = F(0). \quad (6.3)$$

- (b) We may use Taylor’s theorem to expand a function in terms of its derivatives at zero. These derivatives allow us to determine the probabilities in our probability generating function,

$$p_k = \frac{1}{k!} \left. \frac{d^k F}{ds^k} \right|_{s=0}. \quad (6.4)$$

- (c) The probability generating function allows us to compute the average or expected value of  $X$  without tedious calculations involving discrete sums:

$$E(X) \equiv \sum_{k=0}^{\infty} k p_k = \left. \frac{dF}{ds} \right|_{s=1}. \quad (6.5)$$

- (d) The probability generating function also allows us to compute the variance of  $X$  with only a little more effort. In particular,

$$\begin{aligned} \left. \frac{d^2 F}{ds^2} \right|_{s=1} &= \sum_{k=0}^{\infty} (k^2 - k) p_k \\ &= E(X^2) - E(X). \end{aligned} \quad (6.6)$$

However,

$$\text{Var}(X) = E(X^2) - E^2(X), \quad (6.7)$$

so that

$$\text{Var}(X) = \left[ \frac{d^2 F}{ds^2} + \frac{dF}{ds} - \left( \frac{dF}{ds} \right)^2 \right]_{s=1}. \quad (6.8)$$

We can, in other words, compute all the probabilities and statistics that we need in a straightforward way with the probability generating function.

*Example:* Poisson distribution

For the Poisson distribution,

$$\Pr \{X = k\} = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, \dots, \quad (6.9)$$

the probability generating function is

$$F(s) = \sum_{k=0}^{\infty} s^k \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} \quad (6.10)$$

$$= e^{-\lambda} e^{\lambda s} = e^{-\lambda(1-s)} .$$

It is now easy to compute

$$E(X) = \left. \frac{dF}{ds} \right|_{s=1} = \left. \lambda e^{-\lambda(1-s)} \right|_{s=1} = \lambda \quad (6.11)$$

and

$$\begin{aligned} \text{Var}(X) &= \left[ \frac{d^2 F}{ds^2} + \frac{dF}{ds} - \left( \frac{dF}{ds} \right)^2 \right]_{s=1} \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda . \end{aligned}$$

To take full advantage of probability generating functions, I will need two other important results.

*Theorem:* Let  $X$  and  $Y$  be two independent random variables with probability mass functions  $p_k$  and  $q_k$  and generating functions  $F(s)$  and  $G(s)$ . For the new random variable  $Z \equiv X + Y$  with probability mass function  $r_k$ , the probability generating function  $H(s)$  is given by

$$H(s) = F(s)G(s) . \quad (6.12)$$

*Proof:*

$$H(s) = \sum_{k=0}^{\infty} r_k s^k \quad (6.13)$$

$$= \sum_{k=0}^{\infty} \left( \sum_{i=0}^k p_i q_{k-i} \right) s^k \quad (6.14)$$

$$= \sum_{i=0}^{\infty} \sum_{k=i}^{\infty} p_i q_{k-i} s^k \quad (6.15)$$

$$= \sum_{i=0}^{\infty} p_i s^i \sum_{k=i}^{\infty} q_{k-i} s^{k-i} \quad (6.16)$$

$$= F(s)G(s) . \quad (6.17)$$

This result is easily extended to the sum of  $n$  random variables: the probability generating function for the sum of  $n$  variables is simply the corresponding  $n$ -fold product of each individual generating function.

*Example:* Binomial distribution

A Bernoulli random variable  $X$  has the probability distribution

$$\Pr \{Z_i = 1\} = p, \quad \Pr \{Z_i = 0\} = 1 - p. \quad (6.18)$$

The expected value for this distribution is

$$E(Z_i) = (1 \cdot p) + [0 \cdot (1 - p)] = p \quad (6.19)$$

and, since

$$E(Z_i^2) = p, \quad (6.20)$$

the variance of this distribution is just

$$\text{Var}(Z_i) = E(Z_i^2) - E^2(Z_i) = p - p^2 = p(1 - p). \quad (6.21)$$

The probability generating function for a Bernoulli random variable is just

$$F(s) = (1 - p) + ps. \quad (6.22)$$

The binomial distribution is the number of successes (occurrences of  $Z_i = 1$ ) in  $n$  independent Bernoulli trials or, equivalently, the sum of  $n$  independent Bernoulli random variables,

$$X_n = Z_1 + \dots + Z_n. \quad (6.23)$$

It follows that the probability generating function for the binomial distribution is just

$$H(s) = [(1 - p) + ps]^n. \quad (6.24)$$

The expected value of the binomial distribution is

$$E(X_n) \equiv \left. \frac{dH}{ds} \right|_{s=1} = n [(1 - p) + ps]^{n-1} p \Big|_{s=1} = np$$

and, since

$$\begin{aligned} \left. \frac{d^2 H}{ds^2} \right|_{s=1} &= n(n-1) [(1 - p) + ps]^{n-2} p^2 \Big|_{s=1} \\ &= n(n-1) p^2, \end{aligned} \quad (6.25)$$

it follows that the variance of the binomial distribution is just

$$\text{Var}(X_n) = \left[ \left. \frac{d^2 H}{ds^2} + \frac{dH}{ds} - \left( \frac{dH}{ds} \right)^2 \right]_{s=1} \quad (6.26)$$

$$= n p (1 - p) .$$

One can also use the probability generating function for a binomial distribution to show that

$$p_k = \binom{n}{k} p^k (1 - p)^{n-k} . \quad (6.27)$$

*Detour:* Random walks

A slight modification of the above example has immediate relevance to random walks. In the simple, asymmetric random walk that we previously discussed, the probability of moving one step to the right was  $p$  and the probability of moving one step to the left was  $q = 1 - p$ . The displacement at the  $i$ th move was thus a random variable,  $Z_i$ , with the probability distribution

$$\Pr \{Z_i = +1\} = p , \quad \Pr \{Z_i = -1\} = q . \quad (6.28)$$

When I first introduced the probability generating function, I assumed that we were only dealing with nonnegative integer values. For the asymmetric random walk, I may instead take

$$F(s) = p s + \frac{q}{s} \quad (6.29)$$

as the probability generating function for the displacement. Since the position after  $n$  steps is just the sum of  $n$  displacements,

$$X_n = Z_1 + \dots + Z_n , \quad (6.30)$$

we may take the  $n$ -fold product

$$H(s) = \left( p s + \frac{q}{s} \right)^n \quad (6.31)$$

as the probability generating function for the position of our particle.

You can now use this probability generating function to easily show that the expected value and the variance of the position, after  $n$  steps, are

$$E(X_n) = n(p - q) \quad (6.32)$$

and

$$\text{Var}(X_n) = 4 n p q . \quad (6.33)$$

Let us now return to a random variable that takes on only nonnegative values. Suppose now that you are adding  $n$  of these random variables, but that  $n$  is itself the outcome of a random experiment.

*Theorem:* Let  $X_N$  be the sum of  $N$  independent, identically distributed random variables,

$$X_N = Z_1 + \cdots + Z_N , \quad (6.34)$$

each with the common probability generating function  $G(s)$ . If  $N$  ( $\geq 0$ ) is an independent random variable with probability generating function  $F(s)$  (so that we are taking the sum of a random number of identical random variables), the probability generating function  $H(s)$  for  $X_N$  is given by

$$H(s) = F(G(s)) . \quad (6.35)$$

*Proof:* Let  $p_n$  be the probability mass function for  $N$ . Consider  $n$  fixed. Then, by our previous theorem, the generating function for the sum of  $n$  independent and identically distributed random variables is  $[G(s)]^n$ . If we now condition on  $n$ , the probability generating function for  $X_N$  is simply

$$\sum_{n=0}^{\infty} p_n [G(s)]^n . \quad (6.36)$$

However, this is just  $F(s)$ , with  $s$  replaced by  $G(s)$ .

We are now ready to talk about branching processes.

## 6.2. Galton–Watson process

Last quarter, I introduced the simple, linear difference equation,

$$N_{t+1} = R_0 N_t , \quad (6.37)$$

as a density-independent model for a population with discrete, nonoverlapping generations and episodic or pulsed reproduction. Back then, we imagined that each individual left precisely  $R_0$  offspring.

Let us now follow the lead of the Reverend H. W. Watson and of Francis Galton (1874), as paraphrased by Harris (1963), and introduce an element of chance into this formula:

Let  $p_0, p_1, p_2, \dots$  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 descendents in the

male line in any given generation ?

Watson and Galton were interested in the extinction of family names. But to solve their problem, we must develop a generation-by-generation description of the growth of an arbitrary population. In particular, we must determine the population's size,  $N_t$ , in each generation  $t$ . The problem is challenging in that  $N_t$  is now a random variable with a discrete parameter (time) and a countable (semi-infinite) state space.

The Galton–Watson process is a simple, discrete branching process. As stated, this process requires

- (a) that we start with a single individual,

$$N_0 = 1; \quad (6.38)$$

- (b) that the number of offspring sired by that individual be a discrete random variable of given distribution,

$$\Pr \{ N_1 = n \} = p_n, \quad \text{with} \quad \sum_{n=0}^{\infty} p_n = 1; \quad (6.39)$$

and

- (c) that the conditional distribution of  $N_{t+1}$ , given  $N_t = n$ , is the sum of  $n$  independent variables, each with the same distribution as  $N_1$ .

Condition (c) makes the Galton–Watson process a Markov chain. This condition would be specious if, say, infertility is genetic, so that fewer siblings meant fewer children. Condition (c) also implies that our process is density-independent and that individuals do not interfere with one another.

Rather than computing each probability of the Galton–Walton process individually, let us instead determine the probability generating function  $F_t(s)$  for each generation  $t$ . By the first of our requirements for a Galton–Walton process, we have that

$$F_0(s) = s. \quad (6.40)$$

This simply reiterates that we are starting with an individual.

Our second requirement was that the number of offspring sired by this individual be a random variable with probability mass function  $p_n$ . The probability generating function for the first generation can, as a result, be written

$$F_1(s) = F(s), \quad (6.41)$$

where  $F(s)$  is the probability generating function for the distribution or mass function  $p_n$ .

How about the next generation? Well, at the end of the first generation we have a random number of individuals. Each individual will die and leave a random number of offspring. We must thus sum a random number of random variables, each with probability generating function  $F(s)$ , to count the second generation. Thus, by our second theorem from the previous section,

$$F_2(s) = F_1[F(s)] = F \circ F(s). \quad (6.42)$$

We can continue. In general, the members of generation  $t$  are the direct descendents of the  $N_{t-1}$  members of generation  $t-1$ . Thus, we can sum  $N_{t-1}$  random independent variables, each with probability generating function  $F(s)$ , to obtain the *forward* equation

$$F_t(s) = F_{t-1}[F(s)]. \quad (6.43)$$

Alternatively, we can view the members of generation  $t$  as the  $t-1$  generation descendents of the  $N_1$  members of the first generation. We may thus sum  $N_1$  random variables, each with probability generating function  $F_{t-1}(s)$ , to obtain the *backward* equation

$$F_t(s) = F[F_{t-1}(s)]. \quad (6.44)$$

Either way,  $F_t(s)$  is the  $t$ -fold composition of  $F(s)$ ,

$$F_t(s) = F_{t-1}[F(s)] = F[F_{t-1}(s)] = F^t(s). \quad (6.45)$$

In principle, this probability generating function allows us to determine the probability distribution of population sizes in each generation. Only rarely, however, does this composition take a simple form. Even so, equation (6.44) can be used to find the moments of  $N_t$  in each generation, in terms of the moments of  $F(s)$ , and to determine the odds of extinction.

The expected value of  $N_t$  is especially easy to compute. We need simply differentiate equation (6.44) at  $s = 1$ :

$$E(N_t) = \left. \frac{dF_t}{ds} \right|_{s=1}. \quad (6.46)$$

Then, by the chain rule,

$$E(N_t) = F' [F_{t-1}(1)] F'_{t-1}(1) \quad (6.47)$$

$$= F'(1) F'_{t-1}(1) \quad (6.48)$$

$$= R_0 F'_{t-1}(1) \quad (6.49)$$

$$= R_0 E(N_{t-1}), \quad (6.50)$$

where

$$R_0 = F'(1) = E(N_1). \quad (6.51)$$

By repeated application of equation (6.50), we quickly determine that

$$E(N_t) = R_0^t, \quad (6.52)$$

so that the population grows, on average, geometrically with a net reproductive rate equal to the mean number of offspring.

The variance of  $N_t$  is slightly more difficult to compute. We know, from the previous section, that

$$\text{Var}(N_t) = \left[ \frac{d^2 F_t}{ds^2} + \frac{dF_t}{ds} - \left( \frac{dF_t}{ds} \right)^2 \right]_{s=1}. \quad (6.53)$$

The first term is clearly the challenge. We have already seen that

$$F_t(s) = F[F_{t-1}(s)], \quad (6.54)$$

so that

$$F'_t(s) = F'[F_{t-1}(s)] F'_{t-1}(s). \quad (6.55)$$

Differentiating once again gives us

$$F''_t(s) = F'[F_{t-1}(s)] F''_{t-1}(s) + F''[F_{t-1}(s)] [F'_{t-1}(s)]^2, \quad (6.56)$$

so that

$$F''_t(1) = F'(1) F''_{t-1}(1) + F''(1) [F'_{t-1}(1)]^2. \quad (6.57)$$

We may use equation (6.52) to write this more cleanly as

$$F''_{t+1}(1) = R_0 F''_t(1) + R_0^{2t} F''(1). \quad (6.58)$$

**Problem 6.1: Linear nonautonomous difference equations**

Show that equation (6.58) has the solution

$$F_t''(1) = \begin{cases} \frac{(R_0^t - 1) R_0^t F''(1)}{R_0 (R_0 - 1)}, & R_0 \neq 1, \\ t F''(1), & R_0 = 1. \end{cases} \quad (6.59)$$

*Hint:* Let  $F_t''(1) = R_0^t u_t$ .

If  $F(s)$  has variance  $\sigma^2$ , equation (6.53) implies that

$$F''(1) = \sigma^2 + R_0 (R_0 - 1). \quad (6.60)$$

By combining equations (6.53), (6.59), and (6.60) we find that

$$\text{Var}(N_t) = \begin{cases} \frac{(R_0^t - 1) R_0^t}{R_0 (R_0 - 1)} \sigma^2, & R_0 \neq 1, \\ \sigma^2 t, & R_0 = 1. \end{cases} \quad (6.61)$$

Higher moments can be found in a similar manner.

Determining the chance that a lineage goes extinct — Galton and Watson's original problem — has a nice geometric solution. The probability of being extinct in generation  $t$  is  $F_t(0)$ . In light of equation (6.44),

$$F_t(0) = F[F_{t-1}(0)] = F^t(0). \quad (6.62)$$

We may determine the chance of being extinct in any generation by recursively iterating the probability generating function with  $s = 0$ . However, this is the same as cobwebbing along probability generating function  $F(s)$  with initial condition zero!

So, what does the probability generating  $F(s)$  look like? Well, since all the probabilities in definition (6.1) are nonnegative,  $F(s)$  (and its derivative) must be a nondecreasing function on the closed interval  $0 \leq s \leq 1$ . Moreover  $F(0) = p_0 \leq 1$  and  $F(1) = 1$ . For the special cases  $p_0 = 0$  or  $p_0 = 1$ , the problem is uninteresting, so I will restrict our attention to  $0 < p_0 < 1$ . The probability generating function  $F(s)$  may now take three, qualitatively different forms, depending on the size of  $R_0 = F'(1)$ .

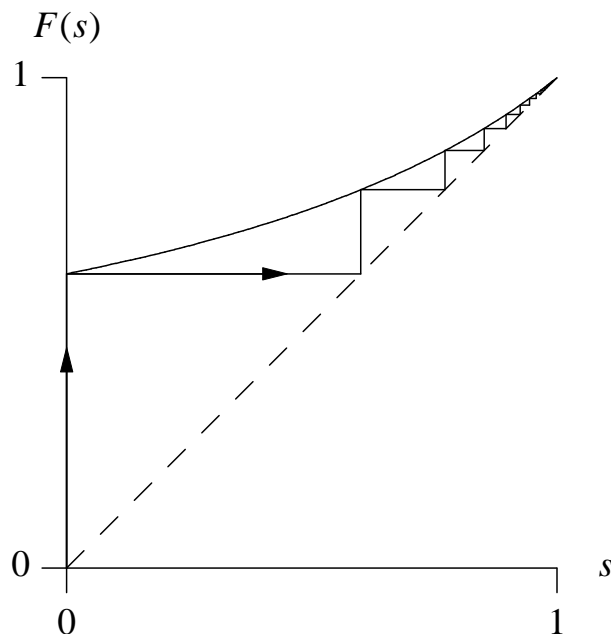


Figure 6.1 Subcritical case

If  $R_0 = F'(1) < 1$  (the subcritical case),  $F(s)$  remains above the  $45^\circ$  line until it finally intersects this line transversally at  $s = 1$  (see Figure 6.1). Cobwebbing now causes  $F_t(0)$  to converge to one: extinction will occur with probability one. If this extinction is slow and gradual, we may look at the conditional distribution for the population size, conditional on no extinction. After one generation, the generating function for this conditional distribution is

$$G(s) = \sum_{n=1}^{\infty} \frac{p_n}{1 - p_0} s^n. \quad (6.63)$$

The denominator may be factored out and the generating function may be rewritten

$$G(s) = \frac{1}{[1 - F(0)]} \sum_{n=1}^{\infty} p_n s^n, \quad (6.64)$$

which simplifies to

$$G(s) = \frac{F(s) - F(0)}{1 - F(0)} \quad (6.65)$$

or

$$G(s) = 1 + \frac{F(s) - 1}{1 - F(0)}. \quad (6.66)$$

By this line of argument, one can also show that

$$G_t(s) = 1 + \frac{F_t(s) - 1}{1 - F_t(0)} \quad (6.67)$$

in generation  $t$ . Thus,  $G_t(s)$  is determined by  $F(s)$ . Moreover, by substituting  $F(s)$  for  $s$ , we obtain

$$G_t[F(s)] = 1 + \frac{F_t[F(s)] - 1}{1 - F_t(0)}, \quad (6.68)$$

which I choose to rewrite as

$$G_t[F(s)] = 1 + \left[ \frac{F_{t+1}(s) - 1}{1 - F_{t+1}(0)} \right] \left[ \frac{1 - F_{t+1}(0)}{1 - F_t(0)} \right], \quad (6.69)$$

or, better yet, as

$$G_t[F(s)] = 1 + [G_{t+1}(s) - 1] \left[ \frac{F_{t+1}(0) - 1}{F_t(0) - 1} \right]. \quad (6.70)$$

For large  $t$ , the second term in square brackets tends to  $R_0$  (consider the linearization of the mapping  $F(s)$  at  $s = 1$ ). The limit of  $G_t(s)$  for large  $t$ ,  $G^*(s)$ , must thus satisfy the functional equation

$$G^*[F(s)] = 1 + [G^*(s) - 1] R_0. \quad (6.71)$$

If a population with a subcritical net reproductive rate has not gone extinct after some long period, it is probably in a “stable” statistically quasistationary state given by this limiting distribution.

If  $R_0 = F'(1) = 1$  (the critical case),  $F(s)$  has both a tangency and a root at  $s = 1$ . Extinction, once again, occurs with probability one, but the approach to extinction is often slow. Indeed, it was shown by Kolmogorov (1938) that if  $R = 1$  and  $F'''(1) < \infty$ , then

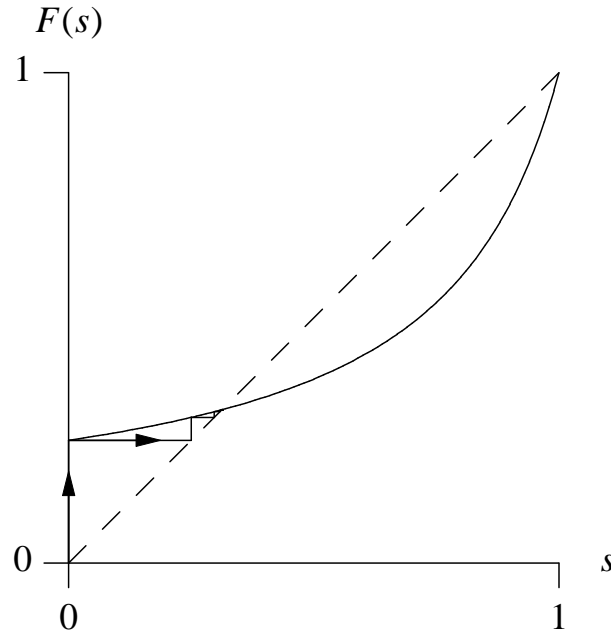
$$P(N_t > 0) \sim \frac{2}{t F''(1)}. \quad (6.72)$$

for large  $t$ . Moreover, despite the certainty of extinction, equations (6.50) and (6.61) imply that

$$E(N_t) = 1, \quad \forall t \geq 1, \quad (6.73)$$

$$\text{Var}(N_t) \rightarrow \infty. \quad (6.74)$$

Hence, if the population has not died out, it may be large.

Figure 6.2  $F(s)$  with  $R_0 > 1$ 

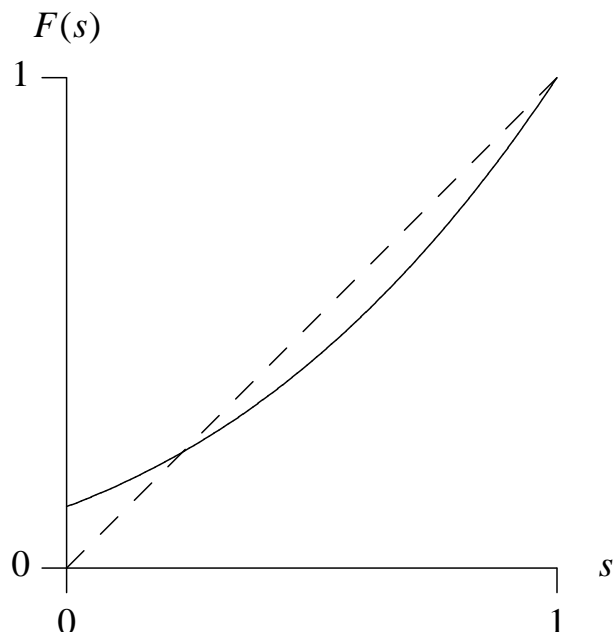
The final alternative is that  $R_0 = F'(1) > 1$  (the supercritical case). Here,

$$F(s) = s \quad (6.75)$$

has a unique solution in the open interval  $(0, 1)$ . The generating function lies above the  $45^\circ$  line to the left of this root and below the  $45^\circ$  line immediately to the right of this root. It rises back up and intersects the  $45^\circ$  line at  $s = 1$  (see Figure 6.2). The fixed point that satisfies equation (6.75) is easily shown to be asymptotically stable. A trajectory that starts at  $s = 0$  rapidly approaches this fixed point. The abscissa of this fixed point is the asymptotic probability of extinction.

*Example:* The case of three children

Suppose that each parent has exactly three children, and that each newborn child is equally likely to be male or female. Let us start with a single male parent ( $N_0 = 1$ ) and let us assume that the number of male progeny can be modeled as a branching process.

Figure 6.3  $F(s)$  for the case of three children

Using a simple binomial distribution, the probabilities for the number of male progeny are

$$p_0 = \frac{1}{8}, p_1 = \frac{3}{8}, p_2 = \frac{3}{8}, p_3 = \frac{1}{8}. \quad (6.76)$$

so that the probability generating function is just

$$F(s) = \frac{1}{8} + \frac{3}{8}s + \frac{3}{8}s^2 + \frac{1}{8}s^3 \quad (6.77)$$

(see Figure 6.3).

The probability of extinction of this male line of descent is just the smaller positive root of  $F(s) = s$ . Thus

$$\frac{1}{8} + \frac{3}{8}s + \frac{3}{8}s^2 + \frac{1}{8}s^3 = s \quad (6.78)$$

$$s^3 + 3s^2 - 5s + 1 = 0 \quad (6.79)$$

$$(s-1)(s^2 + 4s - 1) = 0. \quad (6.80)$$

The probability of extinction is

$$s = -2 + \sqrt{5} = 0.236. \quad (6.81)$$

**Problem 6.2: The survival of right whales (Caswell et al., 1999)**

A female right whale may produce 0, 1, or 2 females the following year. A female at time  $t$  produces 0 offspring if she dies before  $t+1$ , 1 offspring (herself) if she survives without reproducing, and 2 offspring (herself and her calf) if she survives and reproduces. Let  $p$  be the survival probability and let  $m$  be probability of producing a female calf. Thus,

$$p_0 = (1 - p), \quad (6.82a)$$

$$p_1 = p(1 - m), \quad (6.82b)$$

$$p_2 = pm. \quad (6.82c)$$

In 1980,  $p$  and  $m$  were estimated to be  $p = 0.99$  and  $m = 0.063$ . By 1999, these parameters had dropped to  $p = 0.94$  and  $m = 0.038$ . Determine the population growth rate and the extinction probability for 1980 and 1999.

**Problem 6.3: The unfortunate demise of a male line of descent**

Using numbers from the census of 1920, Lotka (1931a, 1931b, 1998) found that the probability mass function for the number of male offspring was well represented by the geometric series

$$p_n = b c^{n-1}, \quad n = 1, 2, \dots, \quad (6.83)$$

with  $b = 0.2126$ ,  $c = 0.5893$ , and  $p_0 = 0.4825$ . Using these figures, determine (a) the probability generating function  $F(s)$ , (b) the net reproductive rate  $R_0$ , and (c) the probability that a newly-created surname will go extinct.

In the supercritical case, if a population does not die out, it diverges to infinity. To better describe this growth, consider the random variable consisting of the population size in the  $t$ th generation normed by the expected population size in generation  $t$ ,

$$W_t \equiv \frac{N_t}{R_0^t}. \quad (6.84)$$

This random variable has several interesting properties. First,

$$E(W_t) = \frac{1}{R_0^t} E(N_t) = 1. \quad (6.85)$$

Secondly,

$$\text{Var}(W_t) = \frac{1}{R_0^{2t}} \text{Var}(N_t) \quad (6.86)$$

so that, for the supercritical case,

$$\text{Var}(W_t) = \frac{\sigma^2}{R_0(R_0 - 1)} \left(1 - \frac{1}{R_0^t}\right). \quad (6.87)$$

Finally, and most importantly, note that

$$E(N_{t+1}|N_t) = R_0 N_t, \quad (6.88)$$

which, by the Markov property, can be rewritten

$$E(N_{t+1}|N_1, N_2, \dots, N_t) = R_0 N_t. \quad (6.89)$$

We may use definition (6.84) to rewrite equation (6.89) in terms of  $W_t$ :

$$E(W_{t+1}|N_1, N_2, \dots, N_t) = W_t. \quad (6.90)$$

This last result is important because it implies that  $W_t$  is a discrete parameter *martingale*.

*Definition:* A sequence  $\{W_t: t \geq 1\}$  is a *martingale* with respect to the sequence  $\{N_t: t \geq 1\}$  if, for all  $t \geq 1$ ,

$$(a) \quad E(|W_t|) < \infty$$

$$(b) \quad E(W_{t+1}|N_1, N_2, \dots, N_t) = W_t.$$

Martingales are important because, subject to minor conditions on the moments of  $W_t$  (satisfied here), they always converge. This is the “martingale convergence theorem” of Doob. A proof may be found in Grimmett and Stirzaker (1992).

The limiting distribution of  $W_t$  is known explicitly in only a few cases. For fractional linear generating functions, the limiting distribution of  $W_t$  is an exponential distribution. In general, the Laplace transform  $\phi(s)$  of the limiting distribution of  $W_t$  satisfies Poincare’s functional equation,

$$\phi(s) = F[\phi(s/R_0)]. \quad (6.91)$$

### 6.3. History of the Galton–Watson process

It was long thought that the theory of branching processes began with the work of Galton (1873) and Watson and Galton (1874). However, it is now appreciated that the French Academician I. J. Bienaymé (1845) anticipated Galton and Watson by some 28 years, as first noted by Heyde and Seneta (1972). Bienaymé was probably stimulated by an empirical study of the duration of noble families by the demographer and statistician L. F. Benoiston de Châteauneuf (1847); he was clearly aware of the importance of the mean number of offspring in determining the probability of extinction. Hence, the Galton–Watson process is increasingly referred to as the Bienaymé–Galton–Watson process. See Kendall (1975) and Heyde and Seneta (1977) for more details.

In 1873, the Swiss botanist Alphonse de Candolle suggested that the extinction of families might have a probabilistic interpretation. Sir Francis Galton (1873) gave Candolle's (1873) suggestion a precise formulation as problem 4001 in the *Educational Times*. However, he received just one answer to his problem — “from a correspondent who totally failed to perceive its intricacy” (Watson and Galton, 1874). He then turned to a friend, the clergyman and mathematician Rev. H. W. Watson, for help. Watson realized that this problem could be solved by iterating probability generating functions, but then, through an algebraic error, incorrectly concluded that *every* surname must die out.

The Galton–Watson process did not reappear in biology until R. A. Fisher (1922, 1930a, 1930b) and J. B. S. Haldane (1927, 1939) used it to study the rate with which rare mutations vanished from a population. The first complete analysis of the probability of extinction was given by the Danish actuary J. F. Steffensen (1930, 1932). Steffensen's solution followed a challenge by A. K. Erlang. Erlang was motivated by the fact that his mother belonged to a well-known but disappearing Danish family (Jagers, 1975); he was unaware of the British work on this problem. A. J. Lotka (1931a, 1931b, 1998) was the first person to compute the probability of extinction using demographic data.

Interest in the Galton–Watson process blossomed after 1940, due largely to its use in modeling nuclear chain reactions (Harris, 1963). Branching processes are now used extensively in biology. One recent book on the topic (Haccou et al., 2005) includes branching process models for the cell cycle, telomere shortening, the polymerase chain reaction (PCR), and measles outbreaks.

#### 6.4. Bellman–Harris process

The Galton–Watson process is a discrete-time branching process in which all individuals live to the same age. In general, semelparous individuals of the same generation may die at different times. The age of reproduction need not be a constant and may itself be a random variable.

Let us imagine that the age of reproduction is a random variable with density function  $g(t)$  and distribution function

$$G(t) = \int_0^t g(u) du, \quad (6.92)$$

that the probability generating function for the number of offspring is

$$H(s) = \sum_{n=0}^{\infty} q_n s^n, \quad (6.93)$$

where  $q_n$  is the probability of  $n$  offspring, and that I may write the probability generating function for the population size  $N(t)$  at time  $t$  as

$$F(t, s) = E[s^{N(t)}] = \sum_{n=0}^{\infty} p_n(t) s^n. \quad (6.94)$$

What is the equation that determines  $F(t, s)$ ?

I will start, as usual, with a single individual. If this individual is alive at time  $t$ , the probability generating function at  $t$  is simply  $s$ . This occurs with probability

$$\int_t^{\infty} g(u) du = 1 - G(t). \quad (6.95)$$

The alternative is that this individual died at some time  $u \leq t$  and was replaced by  $n$  offspring. This occurred with probability  $q_n \cdot g(u) du$ , in which case the probability generating function at time  $t$  is now the sum of  $n$  independent copies of  $F(t - u, s)$ . Adding together the possibilities yields

$$F(t, s) = s [1 - G(t)] + \int_0^t \sum_{n=0}^{\infty} q_n [F(t - u, s)]^n g(u) du \quad (6.96)$$

or

$$F(t, s) = s [1 - G(t)] + \int_0^t H[F(t - u, s)] g(u) du. \quad (6.97)$$

This process was first described by Bellman and Harris (1948, 1952) and is termed an *age-dependent branching* (or Bellman–Harris) process. If the density function for the age of reproduction is exponentially distributed, the Bellman–Harris process is a Markov process. It is non-Markovian for all other densities.

**Problem 6.4: The exponential case**

Let the density function for the age of reproduction be exponentially distributed:

$$g(t) = \lambda e^{-\lambda t}, \quad t \geq 0. \quad (6.98)$$

- (a) Show that equation (6.97) reduces to the differential equation

$$\frac{\partial F}{\partial t} = \lambda H[F(t, s)] - \lambda F(t, s). \quad (6.99)$$

- (b) Solve this differential equation for binary fission,  $H(s) = s^2$ . What is the appropriate initial condition? Determine the probability mass function for the population size  $N(t)$ .

**Problem 6.5: A renewal equation for the mean**

Let

$$M(t) \equiv E[N(t)] \quad (6.100)$$

be the expected value of the population size at time  $t$ . Show that  $M(t)$  satisfies

$$M(t) = 1 - G(t) + R_0 \int_0^t M(t-u) g(u) du. \quad (6.101)$$

Equation (6.101) is an integral equation of the *renewal type*.

**Recommended Readings**

Harris (1963) is the classic reference on branching processes. This book was reprinted in 1989 as a Dover paperback. Bharucha-Reid (1997) covers branching processes within the broader context of Markov processes.

Jagers (1975), Kimmel and Axelrod (2002), and Haccou et al. (2005) consider numerous biologic applications. Moran (1962) discusses some of the early uses of branching processes in population genetics and Kendall (1975) provides a useful history of branching processes. Vatutin and Zubkov (1987, 1993) provide a useful survey of recent results.

## References

- Bellman, R. and Harris, T. 1948. On the theory of age-dependent stochastic branching processes. *Proceedings of the National Academy of Sciences of the United States of America*, **34**, 601–604.
- Bellman, R. and Harris, T. 1952. On age-dependent binary branching processes. *Annals of Mathematics*, **55**, 280–295.
- Benoiston de Châteauneuf, L. F. 1847. Mémoire sur la durée des familles nobles de France. *Memoires de l'Académie royale des sciences morales et politique de l'Institut de France*, **5**, 753–794.
- Bharucha-Reid, A. T. 1997. *Elements of the Theory of Markov Processes and Their Applications*. Dover Publications, Mineola, New York, USA.
- Bienaymé, I. J. 1845. De la loi de multiplacation et de la durée des familles. *Société Philomatique de Paris Extraits*, **5**, 37–39.
- Candolle, A. de. 1873. *Histoire des Sciences et des Savants depuis Deux Siècles*. H. Georg, Geneva, Switzerland.
- Caswell, H., Fujiwara, M., and Brault, S. 1999. Declining survival probability threatens the North Atlantic right whale. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 3308–3313.
- Fisher, R. A. 1922. On the dominance ratio. *Proceedings of the Royal Society of Edinburgh*, **42**, 321–341.
- Fisher, R. A. 1930a. The distribution of gene ratios for rare mutations. *Proceedings of the Royal Society of Edinburgh*, **50**, 204–219.
- Fisher, R. A. 1930b. *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford, UK.
- Galton, F. 1873. *Educational Times*, **19**, 103–105.
- Grimmett, G. R. and Stirzaker, D. R. 1992. *Probability and Random Processes*. Oxford University Press, Oxford, UK.
- Haccou, P., Jagers, P., and Vatutin, V. A. 2005. *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University

- Press, Cambridge, UK.
- Haldane, J. B. S. 1927. A mathematical theory of natural and artificial selection, part V: Selection and mutation. *Proceedings of the Cambridge Philosophical Society*, **23**, 838–844.
- Haldane, J. B. S. 1939. The equilibrium between mutation and random extinction. *Annals of Eugenics*, **9**, 400–405.
- Harris, T. E. 1963. *The Theory of Branching Processes*. Springer-Verlag, New York, New York, USA.
- Heyde, C. C. and Seneta, E. 1972. The simple branching process, a turning point test and a fundamental inequality: a historical note on I. J. Bienaymé. *Biometrika*, **59**, 680–683.
- Heyde, C. C. and Seneta, E. 1977. *I. J. Bienaymé: Statistical Theory Anticipated*. Springer-Verlag, New York, New York, USA.
- Jagers, P. 1975. *Branching Processes with Biological Applications*. John Wiley & Sons, New York, New York, USA.
- Kendall, D. G. 1975. The genealogy of genealogy: branching processes before (and after) 1873. *Bulletin of the London Mathematical Society*, **7**, 225–253.
- Kimmel, M. and Axelrod, D. E. 2002. *Branching Processes in Biology*. Springer, New York, New York, USA.
- Kolmogorov, A. 1938. Zur Lösung einer biologischen Aufgabe. *Izvestiya nauchno-issledovatel'skogo instituta matematiki i mekhaniki pri Tomskom Gosudarstvennom Universitete*, **2**, 1-6.
- Lotka, A. J. 1931a. The extinction of families — I. *Journal of the Washington Academy of Sciences*, **21**, 377–380.
- Lotka, A. J. 1931b. The extinction of families — II. *Journal of the Washington Academy of Sciences*, **21**, 453–459.
- Lotka, A. J. 1998. *Analytical Theory of Biological Populations*. Plenum Press, New York, New York, USA.
- Moran, P. A. P. 1962. *The Statistical Processes of Evolutionary Theory*. Clarendon Press, Oxford, UK.
- Steffensen, J. F. 1930. Om sandsynligheden for at afkommet uddør. *Matematisk Tidsskrift B*, **1**, 19–23.
- Steffensen, J. F. 1932. Deux problèmes du calcul des probabilités. *Annales de l'Institut Henri Poincaré*, **3**, 319–344.

- Vatutin, V. A. and Zubkov, A. M. 1987. Branching processes I. *Journal of Soviet Mathematics*, **39**, 2431–2475.
- Vatutin, V. A. and Zubkov, A. M. 1993. Branching processes II. *Journal of Soviet Mathematics*, **67**, 3407–3485.
- Watson, H. W. and Galton, F. 1874. On the probability of extinction of families. *Journal of the Royal Anthropological Institute*, **4**, 138–144.