

## CHAPTER THREE

---

# The Genealogy of the Birth, Death, and Immigration Process

Simon Tavaré

### 3.1. Introduction

**I**t is indeed a pleasure for me to contribute to this dedicatory volume for Professor Samuel Karlin. Among Karlin's many contributions that address mathematical or statistical issues in the broad area of biology is a collection devoted to the analysis of a variety of stochastic processes that arise in the mathematical theory of population genetics. This theory is the most developed (and the most elegant) in the setting of the infinitely many neutral alleles models, and it is to such problems that this paper is addressed.

In a seminal paper, Karlin and McGregor (1967) describe the following model. Imagine families of individuals each initiated by a single individual at the time points of a renewal process. The size of each family fluctuates in time according to the probabilistic laws of a given stochastic process, different families evolving independently of each other. If we also interpret the families as novel mutant alleles, the model then describes the evolution of the genetic composition of a population of varying size. If all families are given distinct allelic labels, the process may be viewed as a version of the infinitely many neutral alleles model, neutrality here corresponding to the fact that each family is assumed to evolve with identical probabilistic structure.

A detailed description of such a process might include a study of the joint distribution of the numbers of families of different sizes at a given time  $t$ , and its asymptotic behavior as  $t$  increases. Not surprisingly, the most explicit results are available when the renewal process is Poisson, and the families fluctuate according to a linear birth-and-death process. The model just described is then a detailed version of the classical birth-and-death process with immigration. This particular model is one of the generating mechanisms for the well-known Ewens Sampling Formula.

This distribution, which has been the object of detailed study, was derived originally by Ewens (1972) in the context of a population model of fixed but large size; see also Karlin and McGregor (1972). In our variable population size setting, it arises as the joint distribution of the family size statistics, conditional on the total population size; see Watterson (1974) and Kendall (1975).

In fact, I will describe a richer class of processes which keep track of the sizes of families (or number of representatives of each distinct allele, in the population genetics setting) in the order of their appearance in the population. For studying many aspects of the age structure of the families, this approach seems to have some advantages over the unlabeled process described earlier. One consequence of this representation is an age-ordered version of the Ewens Sampling Formula that is a variable population-size analogue of an earlier result of Donnelly and Tavaré (1986).

In Section 3.2 are recorded the basic properties of the birth, death, and immigration process. Section 3.3 provides the probabilistic structure of the age-ordered family size process, and details the connection with size-biasing. Section 3.4 studies distributions of this process that are conditional on the total population size, and describes some connections with reversibility. The final section specializes to the case of the birth process with immigration. Its jump chain is intimately connected to the genealogical structure of the infinitely many neutral alleles model; see Watterson (1984), Hoppe (1984), Donnelly (1986). Its representation via the birth process with immigration provides a simple way to study its asymptotic properties.

### 3.2. The BDI Process

We begin with a brief discussion of the simple (linear) birth-and-death (BD) process. This is a time-homogeneous Markov process whose states are labeled by  $0, 1, 2, \dots$ , and whose behavior is specified by two non-negative parameters:

$\lambda$ , the birth rate per head per unit time

$\mu$ , the death rate per head per unit time.

The (stable, conservative, regular)  $Q$ -matrix of the process has elements determined by

$$\begin{aligned} q_{i,i+1} &= i\lambda, & i &= 0, 1, 2, \dots \\ q_{i,i-1} &= i\mu, & i &= 1, 2, 3, \dots, \end{aligned}$$

the other off-diagonal elements being zero. The number of individuals  $N(t)$  alive in the population at time  $t$  has a well-known distribution (cf. Kendall 1949), and we recall here that

$$P[N(t) = n | N(0) = 1] \equiv g_n(t) = \begin{cases} (1 - a_t)(1 - b_t)b_t^{n-1}, & n \geq 1 \\ a_t, & n = 0, \end{cases} \quad (3.1a)$$

where

$$b_t = \frac{\lambda \exp[(\lambda - \mu)t] - \lambda}{\lambda \exp[(\lambda - \mu)t] - \mu}$$

$$a_t = \mu b_t / \lambda, \quad \text{if } \lambda \neq \mu \quad (3.1b)$$

and

$$a_t = b_t = \frac{\lambda t}{1 + \lambda t}, \quad \text{if } \lambda = \mu. \quad (3.1c)$$

We will be interested in the birth-and-death process with immigration (BDI), which can be constructed from a sequence of independent BD processes as follows. Initially, we suppose there are no individuals in the population. At the time points  $T_1, T_2, \dots$  of a homogeneous Poisson process of rate  $\theta$ , we initiate immigrant families, each starting from a single individual, and evolving independently as a BD process. As the BDI process evolves, families appear, fluctuate in size, and possibly become extinct. Often, we will not distinguish among families of the same size, and so most useful information is contained in the process

$$\{[\xi_0(t), \xi_1(t), \dots], t \geq 0\},$$

in which  $\xi_n(t)$  is the number of families which have  $n$  members at time  $t$ ;  $n = 1, 2, \dots$ , and  $\xi_0(t)$  is the number of extinct families at time  $t$  if  $t \geq T_1$ , and 0 if  $t < T_1$ . Much of the stochastic structure of  $\xi(t)$  is contained in Theorem 1 below.

**THEOREM 1** (Karlin and McGregor 1967):

$$E \prod_{n=0}^{\infty} s_n^{\xi_n(t)} = \prod_{n=0}^{\infty} \exp \left\{ -\theta(1 - s_n) \int_0^t g_n(u) du \right\},$$

for  $|s_n| \leq 1, n = 0, 1, 2, \dots$ . That is, the  $\xi_n(t)$  are independent Poisson-distributed random variables with means

$$E\xi_n(t) = \theta \int_0^t g_n(u) du.$$

It is convenient in the following to scale time so that  $\lambda = 1$ . Then the immigration rate  $\theta$  and the death rate  $\mu$  are the only free parameters, and it follows from (3.1) that

$$E\xi_n(t) = \theta b_i^n/n, \quad n = 1, 2, \dots$$

Hence from THEOREM 1 we obtain

$$P[\xi_i(t) = m_i, i = 1, 2, \dots] = (1 - b_i)^\theta b_i^{\sum r m_r} \prod_{r=1}^{\infty} \frac{1}{m_r!} \left(\frac{\theta}{r}\right)^{m_r}. \quad (3.2)$$

There are a number of other statistics of interest in a BDI process. We mention

$$M(t) = \sum_{j=0}^{\infty} \xi_j(t), \text{ the total number of immigrations up to time } t;$$

$$F(t) = \sum_{j=1}^{\infty} \xi_j(t), \text{ the total number of families alive at time } t;$$

$$I(t) = \sum_{j=1}^{\infty} j\xi_j(t), \text{ the total number of individuals alive at time } t.$$

It follows directly from THEOREM 1 that

$$Es^{F(t)}w^{I(t)} = (1 - b_i)^\theta(1 - wb_i)^{-\theta s},$$

so that  $I(t)$  has a negative binomial distribution with

$$P[I(t) = n] = \binom{\theta + n - 1}{n} (1 - b_i)^\theta b_i^n, \quad n = 0, 1, \dots, \quad (3.3)$$

while  $F(t)$  has a Poisson distribution with mean  $-\theta \ln(1 - b_i)$ .

Further properties of the random variables  $\{\xi_n(t)\}$  may be found in the papers of Karlin and McGregor (1967), Watterson (1974), and Kendall (1975). Kendall's work focuses on the reversibility of the process  $\xi(\cdot) \equiv \{[\xi_1(t), \xi_2(t), \dots], t \geq 0\}$  in the case  $\mu > 1$ ; see Section 3.4 for further details. We turn first, though, to a study of a related process that studies some aspects of the age structure of the families.

### 3.3. An Age-Ordered BDI Process

In this section we extend the analysis of the process  $\xi(t)$  of Section 3.2 by keeping track of the size of extant families *in the order of their appearance*. Recalling that new families arise at the points  $T_1, T_2, \dots$  of a Poisson process of rate  $\theta$ , we say that a family that originated at time  $T_r$  with at

least one member alive at time  $t$  is older than a family extant at time  $t$  that originated at time  $T_s$  if  $r < s$ . We will also keep track of the number  $F(t)$  of families that survive at time  $t$ .

The states  $z$  of our new time-homogeneous Markov process  $\{A(t), t \geq 0\}$  will be either of the form

$z = (0)$ , if no families survive at time  $t$  [i.e.,  $F(t) = 0$ ]; or

$z = (l; \mu_1, \dots, \mu_l)$ , if  $l$  families survive at time  $t$ , and the oldest family has  $\mu_1$  members, the second oldest  $\mu_2$  members,  $\dots$ , the youngest  $\mu_l$  members. It is implicit that  $\mu_1 \geq 1, \dots, \mu_l \geq 1$  for such a  $z$ .

It is a simple matter to compute the transition rates  $q_{zw}$  from state  $z$  to state  $w$ :

#### Immigrations

- (a) If  $z = (l; \eta_1, \dots, \eta_l)$ ,  $w = (l + 1; \eta_1, \dots, \eta_l, 1)$ , and  $l \geq 1$ , then  $q_{zw} = \theta$ .  
 (b) If  $z = (0)$ , and  $w = (1; 1)$ , then  $q_{zw} = \theta$ .

#### Births

If  $z = (l; \eta_1, \dots, \eta_l)$ ,  $w = (l; \eta_1, \dots, \eta_{i-1}, \eta_i + 1, \eta_{i+1}, \dots, \eta_l)$ , and  $l \geq 1$ , then  $q_{zw} = \eta_i \lambda$ ,  $i = 1, \dots, l$ .

#### Deaths

- (a) If  $z = (l; \eta_1, \dots, \eta_l)$ ,  $w = (l; \eta_1, \dots, \eta_{i-1}, \eta_i - 1, \eta_{i+1}, \dots, \eta_l)$ , and  $\eta_i > 1$ , then  $q_{zw} = \eta_i \mu$ .  
 (b) If  $z = (l; \eta_1, \dots, \eta_{i-1}, 1, \eta_{i+1}, \dots, \eta_l)$ ,  
 $w = (l - 1; \eta_1, \dots, \eta_{i-1}, \eta_{i+1}, \dots, \eta_l)$ , then  $q_{zw} = \mu$ .  
 (c) If  $z = (1; 1)$ ,  $w = (0)$ , then  $q_{zw} = \mu$ .

In this treatment we are not interested in keeping track of families that have existed but are now extinct, although this may be of interest elsewhere. We also note that off-diagonal elements of the  $Q$ -matrix not specified above are zero, and the diagonal elements are determined by the requirement that the process be conservative. It also follows by an obvious modification of Kendall's argument (1975, p. 336) that this  $Q$ -matrix is also regular; it is clearly irreducible.

What is interesting is that the distribution of  $A(t)$  can be found explicitly.

**THEOREM 2** *With time standardized so that  $\lambda = 1$ , and  $A(0) = (0)$ ,*

$$P[A(t) = (0)] = (1 - b_t)^{\theta}$$

while for  $l \geq 1, \eta_1 \geq 1, \dots, \eta_l \geq 1,$

$$P[A(t) = (l; \eta_1, \dots, \eta_l)] = \frac{(1 - b_l)^\theta b_l \sum \eta_i \theta^i}{\eta_l(\eta_l + \eta_{l-1}) \cdots (\eta_l + \cdots + \eta_1)}. \tag{3.4}$$

*Proof:* The  $Q$ -matrix of  $A(\cdot)$  is conservative, stable, and regular, so that the forward equations have a unique solution that specifies a probability distribution. We will therefore verify the equations

$$\frac{d}{dt} p_w(t) = \sum_z p_z(t) q_{zw}, \tag{3.5}$$

where  $p_w(t) = P[A(t) = w]$ . Suppose first that  $w = (l; \eta_1, \dots, \eta_l)$  and  $l > 0$ . The left-hand side of (3.5) is

$$(b_l \mu - 1)(\theta b_l + r(1 - b_l)) b_l^{-1} p_w(t), \tag{3.6}$$

where  $\eta_1 + \cdots + \eta_l = r$ . Now consider all the paths that lead from  $z$  to  $w$  in a single change of state. There are two cases to consider.

CASE 1:  $\mu_l > 1$

The possible transitions are

(a)  $z = (l + 1; \eta_1, \dots, \eta_i, 1, \eta_{i+1}, \dots, \eta_l)$ , for which  $q_{zw} = \mu, i = 1, \dots, l,$

and

$z = (l + 1; 1, \eta_1, \dots, \eta_l)$ , for which  $q_{zw} = \mu.$

(b)  $z = (l; \eta_1, \dots, \eta_i + 1, \dots, \eta_l)$ , for which  $q_{zw} = (\eta_i + 1)\mu, i = 1, \dots, l,$

(c)  $z = (l; \eta_1, \dots, \eta_i - 1, \dots, \eta_l)$ , for which  $q_{zw} = (\eta_i - 1), i = 1, \dots, l.$

Using (3.4), we see that

$$\begin{aligned} & \sum_{z \neq w} p_z(t) q_{zw} \\ &= (1 - b_l)^\theta b_l^{l+1} \theta^{l+1} \mu \left\{ \frac{1}{(1 + \eta_l)(1 + \eta_l + \eta_{l-1}) \cdots (1 + \eta_l + \cdots + \eta_1)} \right. \\ & \quad \left. + \frac{1}{\eta_l(\eta_l + 1)(\eta_l + \eta_{l-1} + 1) \cdots (1 + \eta_l + \cdots + \eta_1)} + \cdots \right. \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{\eta_l(\eta_l + \eta_{l-1}) \cdots (\eta_l + \cdots + \eta_1)(1 + \eta_l + \cdots + \eta_1)} \Big\} \\
& + (1 - b_i)^{\theta} b_i^{r+1} \theta^l \mu \times \left\{ \sum_{i=1}^l \frac{1}{\eta_l(\eta_l + \eta_{l-1}) \cdots (\eta_l + \cdots + \eta_{i+1})} \right. \\
& \quad \left. \times (\eta_l + \cdots + \eta_i + 1) \cdots (\eta_l + \cdots + \eta_1 + 1) \right\} \\
& + (1 - b_i)^{\theta} b_i^{r-1} \theta^l \times \left\{ \sum_{i=1}^l \frac{\eta_i - 1}{\eta_l(\eta_l + \eta_{l-1}) \cdots (\eta_l + \cdots + \eta_{i+1})} \right. \\
& \quad \left. \times (\eta_l + \cdots + \eta_i - 1) \cdots (\eta_l + \cdots + \eta_1 - 1) \right\}.
\end{aligned}$$

A little algebra reduces this to

$$\begin{aligned}
& = (1 - b_i)^{\theta} b_i^{r+1} \theta^{l+1} \mu \times \frac{1}{\eta_l(\eta_l + \eta_{l-1}) \cdots (\eta_l + \cdots + \eta_1)} \\
& + (1 - b_i)^{\theta} b_i^{r+1} \theta^l \mu \times \frac{r}{\eta_l(\eta_l + \eta_{l-1}) \cdots (\eta_l + \cdots + \eta_1)} \\
& + (1 - b_i)^{\theta} b_i^{r-1} \theta^l \times \frac{r}{\eta_l(\eta_l + \eta_{l-1}) \cdots (\eta_l + \cdots + \eta_1)} \\
& = p_w(t) [\mu b_i(\theta + r) + r/b_i].
\end{aligned}$$

Since  $q_{ww} = -r(1 + \mu) - \theta$ , the right-hand side of (3.5) becomes

$$\begin{aligned}
& p_w(t) [\mu b_i(\theta + r) + r/b_i - r(1 + \mu) - \theta] \\
& = p_w(t) b_i^{-1} [\mu b_i^2(\theta + r) + r - r b_i(1 + \mu) - \theta b_i] \\
& = p_w(t) b_i^{-1} (\theta b_i + r(1 - b_i))(b_i \mu - 1) \\
& = \text{left-hand side of (3.5), and we are done.}
\end{aligned}$$

#### CASE 2: $\mu_l = 1$

In this case the possible states and their transition rates are given by (a) and (b) above, but (c) must be replaced by

$$\text{(c')} \quad z = (l; \eta_1, \dots, \eta_i - 1, \dots, \eta_{l-1}, 1), \text{ for which } q_{zw} = (\eta_i - 1), \\
i = 1, \dots, l - 1,$$

and

$$z = (l - 1; \eta_1, \dots, \eta_{l-1}), \text{ for which } q_{zw} = \theta.$$

But notice that

$$\sum_{z \text{ of type (c')}} p_z(t) q_{zw}$$

$$\begin{aligned}
&= (1 - b_i)^{\theta} b_i^{r-1} \theta^l \sum_{i=1}^{l-1} \frac{\eta_i - 1}{1(1 + \eta_{l-1}) \cdots (1 + \eta_{l-1} + \cdots + \eta_{i+1})} \\
&\quad \times (\eta_{l-1} + \cdots + \eta_i) \cdots (\eta_{l-1} + \cdots + \eta_1) \\
&\quad + (1 - b_i)^{\theta} b_i^{r-1} \theta^{l-1} \times \frac{1}{\eta_{l-1}(\eta_{l-1} + \eta_{l-2}) \cdots (\eta_{l-1} + \cdots + \eta_1)} \times \theta \\
&= (1 - b_i)^{\theta} b_i^{r-1} \theta^l \times \left\{ \frac{1}{\eta_{l-1}(\eta_{l-1} + \eta_{l-2}) \cdots (\eta_{l-1} + \cdots + \eta_1)} \right. \\
&\quad + \frac{\eta_{l-1} - 1}{\eta_{l-1}(\eta_{l-1} + \eta_{l-2}) \cdots (\eta_{l-1} + \cdots + \eta_1)} \\
&\quad + \frac{\eta_{l-2} - 1}{(1 + \eta_{l-1})(\eta_{l-1} + \eta_{l-2}) \cdots (\eta_{l-1} + \cdots + \eta_1)} + \cdots \\
&\quad \left. + \frac{\eta_1 - 1}{(1 + \eta_{l-1})(1 + \eta_{l-1} + \eta_{l-2}) \cdots (1 + \eta_{l-1} + \cdots + \eta_2)(\eta_{l-1} + \cdots + \eta_1)} \right\} \\
&= (1 - b_i)^{\theta} b_i^{r-1} \theta^l \times \left\{ \frac{1}{(\eta_{l-1} + \eta_{l-2}) \cdots (\eta_{l-1} + \cdots + \eta_1)} \right. \\
&\quad + \frac{\eta_{l-2} - 1}{(1 + \eta_{l-1})(\eta_{l-1} + \eta_{l-2}) \cdots (\eta_{l-1} + \cdots + \eta_1)} + \cdots \\
&\quad \left. + \frac{\eta_1 - 1}{(1 + \eta_{l-1})(1 + \eta_{l-1} + \eta_{l-2}) \cdots (\eta_{l-1} + \cdots + \eta_1)} \right\} \\
&= (1 - b_i)^{\theta} b_i^{r-1} \theta^l \\
&\quad \times \left\{ \frac{1}{(1 + \eta_{l-1})(\eta_{l-1} + \eta_{l-2} + \eta_{l-3}) \cdots (\eta_{l-1} + \cdots + \eta_1)} + \cdots \right. \\
&\quad \left. + \frac{1}{(1 + \eta_{l-1})(1 + \eta_{l-1} + \eta_{l-2}) \cdots (\eta_{l-1} + \cdots + \eta_1)} \right\} = \cdots \\
&= (1 - b_i)^{\theta} b_i^{r-1} \theta^l \times \frac{r}{(1 + \eta_{l-1})(1 + \eta_{l-1} + \eta_{l-2}) \cdots (1 + \eta_{l-1} + \cdots + \eta_1)} \\
&= p_w(t)r/b_i.
\end{aligned}$$

The remainder of the verification of (3.5) now proceeds just as in Case 1. The proof may be completed by checking the remaining case, which occurs when  $l = 0$ . This is easy, and the details are omitted.

It is worthwhile at this stage to comment on the relationship between the distribution of  $A(t)$  specified by (3.4) and that of  $\xi(t)$  given by (3.2).



Assume then that for a given  $t > 0$ , the extant family sizes  $\eta_1, \dots, \eta_l$  correspond to family size statistics  $m_1, m_2, \dots$  given by

$$m_i = \text{card}\{j: \eta_j = i\}, \quad i = 1, 2, \dots, \tag{3.7a}$$

satisfying

$$m_1 + m_2 + \dots = l, \tag{3.7b}$$

and

$$m_1 + 2m_2 + 3m_3 + \dots = \eta_1 + \eta_2 + \dots + \eta_l. \tag{3.7c}$$

The distribution of  $\xi(t)$  follows from that of  $A(t)$  by a collapsibility argument. For  $m_1, m_2, \dots$  satisfying the conditions of (3.7), we see that

$$P[\xi_i(t) = m_i, i = 1, 2, \dots] = \sum P[A(t) = (l; \eta_1, \eta_2, \dots, \eta_l)],$$

the sum ranging over all  $\eta_1, \dots, \eta_l$  satisfying (3.7a). But this last is

$$\left( \prod_{r \geq 1} m_r! \right)^{-1} \sum P[A(t) = (l; \eta_{\pi(1)}, \dots, \eta_{\pi(l)})],$$

where  $\pi$  denotes a permutation of  $\{1, 2, \dots, l\}$ , and the combinatorial quantity arises because the sum over all permutations  $\pi$  counts the required terms  $\prod m_r!$  times. Observing that

$$\sum \frac{1}{\eta_{\pi(l)}(\eta_{\pi(l)} + \eta_{\pi(l-1)}) \dots (\eta_{\pi(1)} + \dots + \eta_{\pi(l)})} = \frac{1}{\eta_1 \eta_2 \dots \eta_l},$$

we obtain

$$\begin{aligned} P[\xi_i(t) = m_i, i \geq 1] &= \frac{(1 - b_t)^\theta \theta^l b_t^\Sigma \eta_i}{\eta_1 \eta_2 \dots \eta_l} \times \frac{1}{\prod_{i \geq 1} m_i!} \\ &= (1 - b_t)^\theta b_t^\Sigma r^{m_r} \prod_{r \geq 1} \left(\frac{\theta}{r}\right)^{m_r} \frac{1}{m_r!}, \end{aligned}$$

this last following because  $\eta_1 \dots \eta_l = 1^{m_1} 2^{m_2} \dots$ , and the relationships in (3.7) hold.

Conversely, there is a simple operational description of how the age-ordering arises from the process  $\xi(\cdot)$ . This relies on size-biasing in the following way. Imagine that  $\xi_i(t) = m_i, i \geq 1$ , that  $I(t) = n$ , and that (3.7) holds. Randomly select an individual, and remove him and all members of his family. These are assigned the label 1. Next, randomly choose one of the remaining individuals, and remove him and all his family. This group is given the label 2. Continue in this way until the remaining family

is labeled  $l$ . This assignment produces  $\eta_1$  individuals with label 1,  $\eta_2$  with label 2,  $\dots$ ,  $\eta_l$  with label  $l$  with conditional probability (given the family sizes  $m_i$ ) of

$$\begin{aligned} \left(\prod_r m_r!\right) \frac{\eta_1}{n} \times \frac{\eta_2}{n - \eta_1} \times \dots \times \frac{\eta_{l-1}}{\eta_l + \eta_{l-1}} \\ = \left(\prod_r m_r!\right) \frac{\eta_1 \eta_2 \dots \eta_l}{n(\eta_2 + \dots + \eta_l) \dots (\eta_l + \eta_{l-1}) \eta_l}. \end{aligned}$$

Hence the unconditional probability of getting  $\eta_1$  individuals with label 1,  $\dots$ ,  $\eta_l$  with label  $l$  is

$$\begin{aligned} (1 - b_l)^\theta b_l^{\sum r m_r} \frac{\theta^{\sum m_r}}{\left(\prod_r m_r!\right) \left(\prod_r r^{m_r}\right)} \left(\prod_r m_r!\right) \\ \times \frac{\eta_1 \dots \eta_l}{\eta_l(\eta_l + \eta_{l-1}) \dots (\eta_l + \dots + \eta_1)}, \end{aligned}$$

which reduces to the distribution in (3.4) after simplification. Thus we see that the joint distribution of the number of individuals in the oldest family,  $\dots$ , the youngest family is the same as that of the number of individuals labeled 1, 2,  $\dots$  in the size-biased permutation.

### 3.4. Conditional Distributions and Reversibility

This section is divided into two parts, the first devoted to the structure of the processes  $A(\cdot)$  and  $\xi(\cdot)$  conditional on the total population size  $I(\cdot)$ , and the second to reversibility.

#### 3.4.1 CONDITIONAL DISTRIBUTIONS

In the sequel, we will let  $P_n(\cdot)$  denote conditional probabilities given that  $I(t) = n$ . It follows from (3.2) and (3.3) that

$$P_n[\xi_i(t) = m_i, i = 1, 2, \dots, n] = \binom{\theta + n - 1}{n}^{-1} \prod_{j=1}^n \binom{\theta}{j}^{m_j} \frac{1}{m_j!}. \quad (3.8)$$

Notice that this distribution is independent of  $t$  (and  $\mu$ ), so that suppression of the  $t$  in  $P_n(\cdot)$  is justified. The distribution (3.8) is known in the population genetics literature as the Ewens Sampling Formula (Ewens 1972). It was derived in the present context by Watterson (1974), and Kendall (1975).

The age-ordered version follows from (3.3) and (3.4):

$$P_n[A(t) = (l; n_1, \dots, n_l)] = \binom{\theta + n - 1}{n}^{-1} \frac{\theta^l}{n_l(n_l + n_{l-1}) \cdots (n_l + \cdots + n_1)}. \quad (3.9)$$

Donnelly and Tavaré (1986) showed that (3.9) arises as the distribution of an age-ordered sample of size  $n$  from a stationary Moran model of constant size, and also in the limit of large population size in a wide variety of other models.

Of course, (3.8) and (3.9) may be related by the same size-biasing argument described in the last section. This same idea can be applied in several other ways also. For example, conditional on  $I(t) = n$ , the probability that a given family of size  $i$  is the oldest is  $i/n$ , since this is the probability that this family of size  $i$  is the first chosen in the size-biased sample. Here is another example.

Let  $O$  denote the number of individuals in the oldest family. Then from (3.9) we obtain

$$P_n[F(t) = l; O = j] = \sum \frac{1}{(\eta_2 + \cdots + \eta_l) \cdots (\eta_{l-1} + \eta_l) \eta_l} \times \frac{\theta^l n!}{\theta_{(n)} n},$$

where  $\theta_{(n)} = \theta(\theta + 1) \cdots (\theta + n - 1)$ , and the summation is over

$\eta_2, \dots, \eta_l \in \{1, 2, \dots, n - j\}$  with  $\eta_2 + \cdots + \eta_l = n - j$ . This last sum is easily effected, revealing that

$$P_n[F(t) = l; O = j] = \frac{\theta n! |S_{n-j}^{(l-1)}|}{n \theta_{(n)} (n - j)!}, \quad (3.10)$$

where  $S_n^{(j)}$  is a Stirling number of the first kind. For given values of  $j$  and  $n$ ,  $F(t)$  may take any value between 1 and  $n - j + 1$ . Hence summing (3.10) over  $l = 1, 2, \dots, n - j + 1$  gives the marginal distribution of  $O$ :

$$P_n[O = j] = \frac{\theta n! \theta_{(n-j)}}{n \theta_{(n)} (n - j)!}; \quad (3.11)$$

See Kelly (1977). By size-biasing, it is clear that

$$P_n[O = j] = \frac{j E[\xi_j(t) | I(t) = n]}{n},$$

whence from (3.11), we obtain

$$E[\xi_j(t) | I(t) = n] = \frac{\theta n! \theta_{(n-j)}}{j \theta_{(n)} (n - j)!},$$

as found by Karlin and McGregor (1967), equation (4.13).

## 3.4.2 LIMIT DISTRIBUTIONS AND REVERSIBILITY

In this section I will describe some of the connections between the age-ordered BDI process and reversibility. Let's assume that the death rate  $\mu$  is greater than the birth rate  $\lambda$  ( $\equiv 1$  in the following). Any family that arises in the BDI must, with probability one, become extinct in finite time. It follows immediately from (3.4) that for any state  $z = (l; \eta_1, \dots, \eta_l)$ ,

$$\lim_{t \rightarrow \infty} P[A(t) = z | A(0) = (0)] \equiv \pi(z) = \frac{\left(1 - \frac{1}{\mu}\right)^\theta \left(\frac{1}{\mu}\right)^{\sum \eta_i} \theta^l}{\eta_l(\eta_l + \eta_{l-1}) \cdots (\eta_l + \cdots + \eta_1)}. \quad (3.12)$$

Since the state space is irreducible, it also follows that

$$\lim_{t \rightarrow \infty} P[A(t) = z | A(0) = w] = \pi(z)$$

for any  $w$ . The distribution  $\{\pi(z)\}$  is also the invariant measure of the process  $A(\cdot)$ , and we can if we wish arrange to extend the definition of the process to the whole time axis  $(-\infty, \infty)$  in such a way that the homogeneity is preserved, and such that  $P[A(t) = z] = \pi(z)$  for all times  $t$  and states  $z$ . Assume this has been done. The process  $A(\cdot)$  provides a rather detailed description of the appearance and disappearance of families through time. Keeping track of the numbers of families of different sizes (rather than their age order) can be achieved by grouping collections of states  $z = (l; \eta_1, \dots, \eta_l)$  of the  $A(\cdot)$  process which have the same family-size statistics  $(m_1, m_2, \dots)$  [cf. (3.7)]. This lumped process is precisely the stationary "FS-process"  $\xi(\cdot)$  discussed so eloquently by Kendall (1975).

Recall that an irreducible process with (regular, stable, and conservative)  $Q$ -matrix  $\{q_{zw}\}$  is symmetrically reversible with respect to the measure  $\pi(\cdot)$  if and only if

$$\pi(z)q_{zw} = \pi(w)q_{wz} \quad \text{for all } z, w. \quad (3.13)$$

Kendall showed *inter alia* that the process  $\xi(\cdot)$  is symmetrically reversible with respect to the measure  $\pi$  determined by letting  $t \rightarrow \infty$  in (3.2). It is tempting to address questions about the ages of families (rather than just their age ordering) by exploiting this reversibility; the past history of a family should be stochastically similar to its future. However, from a realization of the family-size process *alone* this is not possible, since from it one cannot follow the progress of a given family from birth to extinction. What is needed is a more sophisticated, labeled process that allows families to be distinguished; cf. Watterson (1976) and Kelly (1979, p. 152) in the context of the Moran model. While the process  $A(\cdot)$  does give some information about age-ordering, there seems to be no direct way to use it

to study the ages themselves. This process is clearly not reversible with respect to the measure  $\pi$  in (3.12) [take, for example,  $z = (2; 2, 2)$  and  $w = (2; 1, 2)$ ; the ratio of the two sides of (3.13) is  $3/4$ ] and its time reversal has a very complicated structure.

### 3.5. The Birth Process with Immigration

One special case of the BDI model is the linear birth process with immigration. The simplicity of the structure of this process allows the development of a much richer asymptotic theory than is presently known for the case allowing death. Details and further ramifications may be found elsewhere (Tavaré 1987). We will content ourselves here with an overview of the results.

Formally, the birth process with immigration is the special case of the BDI in which the death rate  $\mu = 0$ . Notice that since death is impossible, a family grows without limit as time increases and, in particular, can never go extinct. Since a Poisson process in reversed time is still a Poisson process (with the same rate,  $\theta$  in this case), all questions about the ages themselves of families in the BDI process can be answered immediately. Here interest focuses on the asymptotic behavior of the sizes of these families as time increases.

Again, we will use the notation  $A(t)$  to denote the value of the (age-ordered) process at time  $t$ , and we assume  $A(0) = (0)$ . We will assume in the sequel that time has been scaled so that the birth rate is  $\lambda = 1$ . From (3.4) with  $\mu = 0$  it follows that

$$P[A(t) = 0] = e^{-\theta t},$$

$$P[A(t) = (l; \eta_1, \dots, \eta_l)] = \frac{e^{-\theta t}(1 - e^{-t})^{\sum \eta_i \theta^l}}{\eta_l(\eta_l + \eta_{l-1}) \cdots (\eta_l + \cdots + \eta_1)}. \quad (3.14)$$

Notice that the first term in (3.14) is just the probability that no events have occurred in the Poisson process in time  $t$ . We will also require the distribution of  $I(t)$ , the total number of individuals in the population at time  $t$  [and  $I(0) = 0$ ]. This follows in the same way from (3.3) as

$$P[I(t) = n] = \binom{\theta + n - 1}{n} e^{-\theta t}(1 - e^{-t})^n, \quad n = 0, 1, \dots \quad (3.15)$$

We will now decompose the structure of the process  $\{A(t), t \geq 0\}$  into its jump chain  $\{J_n, n \geq 0\}$  and its time-scale process. To this end, define  $\tau_0 = 0$ , and let  $\tau_n$  be the time of the  $n$ th change of state of the process. It will be convenient to let

$$S_n = \{(l; \mu_1, \dots, \mu_l) : \mu_1 + \cdots + \mu_l = n\}$$

be the subset of the state space of  $A(\cdot)$  with a total of  $n$  individuals, for  $n = 0, 1, 2, \dots$ . The jump chain  $J_0 = (0)$ ,  $J_n = A(\tau_n +)$ ,  $n = 1, 2, \dots$  has one-step transition probabilities determined by

$$P[J_n = (l; \eta_1, \dots, \eta_{i-1}, \eta_i + 1, \eta_{i+1}, \dots, \eta_l) | J_{n-1} = (l; \eta_1, \dots, \eta_l)] = \frac{\eta_i}{n - 1 + \theta}, \quad i = 1, 2, \dots, l. \tag{3.16}$$

$$P[J_n = (l + 1; \eta_1, \dots, \eta_l, 1) | J_{n-1} = (l; \eta_1, \dots, \eta_l)] = \frac{\theta}{n - 1 + \theta},$$

if  $(l; \eta_1, \dots, \eta_l) \in S_{n-1}$ .

It can be shown that  $\{J_n, n \geq 0\}$  and  $\{I(t), t \geq 0\}$  are independent stochastic processes, and since  $A(t) = J_{I(t)}$ ,  $t \geq 0$ , it follows that

$$P[A(t) = (l; \eta_1, \dots, \eta_l)] = \sum_{n \geq 0} P[J_n = (l; \eta_1, \dots, \eta_l)] P[I(t) = n].$$

Hence,

$$P[J_n = (l; \eta_1, \dots, \eta_l)] = \binom{\theta + n - 1}{n}^{-1} \frac{\theta^l}{\eta_l(\eta_l + \eta_{l-1}) \cdots (\eta_l + \cdots + \eta_1)} \tag{3.17}$$

if  $(l; \eta_1, \dots, \eta_l) \in S_n$ .

The jump chain  $\{J_n, n \geq 0\}$  is a Markov chain that arises in the study of the genealogy of the stationary infinitely many neutral alleles model. Hoppe (1984) and Watterson (1984) describe a Pólya-like urn with a transition mechanism similar to that of  $\{J_n, n = 0, 1, \dots\}$ . They focus on the fact that the urn model gives rise to the Ewens sampling formula. Connections between their process and results on age ordering are developed and exploited in Donnelly (1986) and Donnelly and Tavaré (1986). See also Aldous (1985) for a related model.

Our attention is now directed to the asymptotic behavior of the family sizes in  $A(t)$  as  $t \rightarrow \infty$ . To study this question recall that the new families arise at the points  $T_1, T_2, \dots$  of a homogeneous Poisson process of rate  $\theta$ . We will change notation slightly, and define

$$\zeta_n(t) = \begin{cases} 0, & \text{if } T_n > t \\ \text{size of family initiated at time } T_n, & \text{if } T_n \leq t. \end{cases}$$

Now recall for a moment how a typical family evolves. A family initiated at time 0, say, with a single individual grows according to a linear birth process  $[N(t), t \geq 0]$  of rate 1. It is a well-known fact about such a process that  $e^{-t}N(t)$  converges almost surely as  $t \rightarrow \infty$  to a random variable having an exponential distribution with mean 1. It follows from this structure

that if  $E_1, E_2, \dots$  are independent and identically distributed random variables having the exponential distribution with mean 1, then

$$e^{-t}[\zeta_1(t), \zeta_2(t), \dots] \longrightarrow (e^{-T_1}E_1, e^{-T_2}E_2, \dots)$$

almost surely as  $t \rightarrow \infty$ .

There are several interesting consequences that flow from this result once the structure of the limit vector is uncovered. As an example, notice that since  $T_1, T_2, \dots$  are the points of a Poisson process of rate  $\theta$ , and the  $\{E_i\}$  are i.i.d., the collection  $\{(T_i, E_i), i = 1, 2, \dots\}$  may be identified as the points of a marked Poisson process. From this it can be shown that the points  $\{E_i \exp(-T_i), i = 1, 2, \dots\}$  may be viewed as the points (in some order) of a Poisson process on  $(0, \infty)$  with mean measure density  $\theta e^{-x}/x, x > 0$ . Hence it follows that, almost surely as  $t \rightarrow \infty$ ,  $I(t)^{-1}[\zeta_1(t), \zeta_2(t), \dots] \rightarrow (P_1, P_2, \dots)$ , say, and a calculation establishes that the random variables  $\{P_i\}$  have the representation

$$P_i = (1 - Z_1)(1 - Z_2) \cdots (1 - Z_{i-1})Z_i, \quad i = 1, 2, \dots, \quad (3.17)$$

where the  $\{Z_i\}$  are independent and identically distributed random variables with density  $\theta(1 - x)^{\theta-1}, x \in (0, 1)$ .  $P_i$  is the (asymptotic) fraction of the population that belongs to the  $i$ th oldest family.

The decreasing-order statistics of the random variables  $\{P_i\}$  have the Poisson-Dirichlet distribution with parameter  $\theta$ ; see Kingman (1975). Such random vectors arise not only in the population genetics setting (see Kingman 1980), but also in the context of species abundance models (McCloskey 1965 and Engen 1978, for example), and (when  $\theta = 1$ ) in the theory of random permutations (Vershik and Shmidt 1977).

### Acknowledgments

I would like to thank Dr. Peter Donnelly and Professor A. G. Pakes for critical reviews of earlier drafts of this paper. This research was supported in part by N.S.F. grant DMS 86-08857.

### References

- Aldous, D. J. 1985. Exchangeability and related topics. *Lecture Notes in Mathematics* 1117: 1–198. Springer-Verlag, New York.
- Donnelly, P. J. 1986. Partition structures, Pólya urns, the Ewens sampling formula and the ages of alleles. *Theor. Pop. Biol.* 30: 271–288.
- Donnelly, P. J., and Tavaré, S. 1986. The ages of alleles and a coalescent. *Adv. Appl. Prob.* 18: 1–19.
- Engen, S. 1978. *Stochastic abundance models*. Halsted Press, New York.

- Ewens, W. J. 1972. The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* 3: 87–112.
- Hoppe, F. M. 1984. Pólya-like urns and the Ewens sampling formula. *J. Math. Biol.* 20: 91–94.
- Karlin, S., and McGregor, J. 1967. The number of mutant forms maintained in a population. In *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability*, ed. L. LeCam and J. Neyman, pp. 415–438. University of California Press, Berkeley.
- Karlin, S., and McGregor, J. 1972. Addendum to a paper of W. Ewens. *Theor. Pop. Biol.* 3: 113–116.
- Kelly, F. P. 1977. Exact results for the Moran neutral allele model. *Adv. Appl. Prob.* 9: 197–201.
- Kelly, F. P. 1979. *Reversibility and Stochastic Networks*. Wiley, New York.
- Kendall, D. G. 1949. Stochastic processes and population growth. *J. Roy. Statist. Soc. B* 11: 230–264.
- Kendall, D. G. 1975. Some problems in mathematical genealogy. In *Perspectives in Probability and Statistics: Papers in honour of M. S. Bartlett*, pp. 325–345. Ed. J. Gani. Applied Probability Trust, Sheffield. Academic Press, London.
- Kingman, J.F.C. 1975. Random discrete distributions. *J. Roy. Statist. Soc. B* 37: 1–22.
- Kingman, J.F.C. 1977. The population structure associated with the Ewens sampling formula. *Theor. Pop. Biol.* 11: 274–283.
- Kingman, J.F.C. 1980. *Mathematics of Genetic Diversity*. CBMS Regional Conference Series in Applied Math. 34. S.I.A.M., New York.
- Kingman, J.F.C. 1982. The coalescent. *Stoch. Proc. Appl.* 13: 235–248.
- McCloskey, J. W. 1965. A model for the distribution of individuals by species in an environment. Ph.D. thesis, Michigan State University.
- Tavaré, S. 1987. The birth process with immigration, and the genealogical structure of large populations. *J. Math. Biol.* 25: 161–168.
- Vershik, A. M., and Shmidt, A. A. 1977. Limit measures arising in the asymptotic theory of symmetric groups. I. *Theor. Prob. Applns.* 22: 70–85.
- Watterson, G. A. 1974. The sampling theory of selectively neutral alleles. *Adv. Appl. Prob.* 6: 463–488.
- Watterson, G. A. 1976. Reversibility and the age of an allele. 1. Moran's infinitely many neutral alleles model. *Theor. Pop. Biol.* 10: 239–253.
- Watterson, G. A. 1984. Estimating the divergence time of two species. *Statistical Research Report* 94, Monash University, Australia.