SIAM REVIEW
Vol. 46, No. 2, pp. 202–229

In the Garden of Branching Processes*

Karin S. Dorman†
Janet S. Sinsheimer‡
Kenneth Lange§

Abstract. The current paper surveys and develops numerical methods for Markovian multitype branching processes in continuous time. Particular attention is paid to the calculation of means, variances, extinction probabilities, and marginal distributions in the presence of a Poisson stream of immigrant particles. The Poisson process assumption allows for temporally complex patterns of immigration and facilitates application of marked Poisson processes and Campbell's formulas. The methods and formulas derived are applied to four models: two population genetics models, a model for vaccination against an infectious disease in a community of households, and a model for the growth of resistant HIV virus in patients undergoing drug therapy.

Key words. birth–death process, branching process, immigration, progeny generating function, matrix exponential, Poisson process, Kronecker product, mutation, X-linked, haplotype, vaccination, HIV

AMS subject classifications. 60J85, 92B05, 92D10, 92D25, 92D30

DOI. 10.1137/S0036144502417843

1. Introduction.

The theory of branching processes has a long and venerable history [1, 2, 11, 14, 20, 21, 22, 32, 38]. Connections with specific applications, such as surname survival and mutant gene extinction, were initially so strong that the basic concepts were rediscovered in several different settings. Inevitably, however, concern with particular applications gave way to rigorous explorations of the probabilistic underpinnings of the subject. Because exact solutions are often very difficult if not impossible to derive, emphasis also shifted to asymptotic results. These beautiful findings and various generalizations of branching processes form the core of the current theory.

In our view, it is time to take stock of the development of branching processes. Theoretical advances have slowed as probabilists move on to more exotic subjects. This is a pity because branching processes have much to offer in applications, particu-

*Received by the editors November 13, 2002; accepted for publication (in revised form) August 27, 2003; published electronically May 3, 2004.
http://www.siam.org/journals/sirev/46-2/41784.html

†Departments of Statistics and Genetics, Development & Cell Biology, Iowa State University, Ames, IA 50011 (kdorman@iastate.edu). This author was supported in part by NIH predoctoral fellowship GM08185.

‡Departments of Biomathematics, Biostatistics, and Human Genetics, University of California, Los Angeles, CA 90095 (janet@mednet.ucla.edu). This author was supported in part by USPHS grants MH59490 and CA16042.

§Departments of Biomathematics, Human Genetics, and Statistics, University of California, Los Angeles, CA 90095 (klange@ucla.edu). This author was supported in part by USPHS grants GM53275 and MH59490.

larly in biological fields such as epidemiology, ecology, evolution, genetics, population biology, and cell biology [20, 22, 28, 42]. Although many mathematical modelers enthusiastically embrace branching processes, they are often thwarted by real difficulties in using the general theory. Most of these difficulties are numerical. Basic problems such as finding the full distribution of the number of current particles appear intractable and discourage application.

In the current paper, we survey and extend existing numerical methods for multitype, continuous-time, Markovian processes. There is little need to argue the need for multitype processes. The case for continuous time as opposed to discrete time is more subtle. One reason favoring continuous-time processes is that they better incorporate differences in particle lifetimes. Another, perhaps less obvious, reason is that continuous-time processes accommodate immigration of particles from the external world more gracefully. Here we have in mind immigration mediated by a Poisson process. Most applications involve immigration of this sort, and the imposition of the Poisson assumption greatly facilitates derivation of explicit formulas. This fact, apparently first recognized by Bartlett [3], has been largely forgotten in modern applications. The key links between Poisson processes and branching processes are coloring, marking, and Campbell's formulas [23].

Finally, we focus on Markovian branching processes, where particle lifetimes are exponentially distributed. The Markovian assumption entails such radical theoretical simplifications that it is hard to argue against its inclusion in most models. Of course, some biological applications involve exponential lifetimes and some do not. The decision to employ age-dependent branching processes has to be made on a case-by-case basis. Fortunately, the recent book of Kimmel and Axelrod [22] provides sound advice on this difficult choice.

The following sections survey existing theory in a nonrigorous manner with stress on concrete formulas and numerical techniques. Some of these techniques, such as our methods for finding marginal distributions and means and variances in the presence of nonhomogeneous immigration rates, are novel. Matrix exponential functions and Kronecker products figure prominently in our derivations. At the conclusion of our theoretical summary, we tackle four biological examples that illustrate the numerical methods in action. Two of these examples are genetic, another concerns vaccination strategies, and the last concerns HIV evolution within a single patient. All four examples rely on the fundamental assumption that particles have independent lifespans and reproductive outcomes. This assumption is both the major strength and weakness of the theory of branching processes. While it excludes interaction between particles, it drives the whole theory and enables all derivations.

2. Definition of a Multitype Branching Process. A multitype branching process records the history of a finite number of independently acting particles that reproduce and die. Each particle is classified in one of r possible categories. In the continuous-time processes that we study, a type-i particle lives an exponentially distributed length of time with death rate ω_i. The "lack of memory" property of the exponential distribution guarantees that a continuous-time process is Markovian. At the end of its life, a particle reproduces both particles of its own type and particles of other types as summarized by the progeny generating function

$$P_i(s) \;=\; \sum_j p_{ij} s^j \;=\; \sum_j p_{ij} s_1^{j_1} \cdots s_r^{j_r}.$$

In this definition, the entries of the row vector $s = (s_1, \ldots, s_r)$ and the multi-index $j = (j_1, \ldots, j_r)$ range over the unit interval $[0, 1]$ and the nonnegative integers, respectively. The coefficient p_{ij} is the probability that a type-i particle gives birth to j_1 particles of type 1, j_2 particles of type 2, and so forth. Extinction is possible provided $p_{i\mathbf{0}} > 0$ for some i.

The random vector $Z_t = (Z_{t1}, \ldots, Z_{tr})$ of particle counts is the primary object of interest in a multitype branching process. In continuous time, the time index t is any nonnegative real number. The process starts at time 0 and evolves thereafter according to strict probabilistic rules. Finding moments and understanding the distribution of Z_t is best done via the probability generating functions

$$Q_i(t, s) = \mathrm{E}\,(s^{Z_t} \mid Z_t = u_i).$$

Here u_i is the unit row vector with all entries 0 except for a 1 at position i. Occasionally we will use the abbreviations

$$\begin{aligned} Q(t, s) &= [Q_1(t, s), \ldots, Q_r(t, s)], \\ P(s) &= [P_1(s), \ldots, P_r(s)]. \end{aligned}$$

The utility of probability generating functions stems from the product formula

$$\mathrm{E}\left(s^{X+Y}\right) = \mathrm{E}\left(s^X\right) \mathrm{E}\left(s^Y\right)$$

for independent random vectors X and Y. This formula extends to sums involving a random number of random vectors. Hence, if $Y = \sum_{i=1}^{N} X_i$, where the X_i are independent and identically distributed random vectors with common multivariate probability generating function $G(s)$ and N is a random integer with univariate probability generating function $F(w)$, then Y is a random vector with probability generating function $\mathrm{E}\left(s^Y\right) = F[G(s)]$.

The qualitative behavior of a multitype branching process is to a large extent governed by the $r \times r$ matrix Ω whose entry in row i and column j is $\omega_i(f_{ij} - 1_{\{i=j\}})$, where $f_{ij} = \frac{\partial}{\partial s_j} P_i(\mathbf{1})$ is the expected number of offspring particles of type j produced by a particle of type i. Because the associated matrix $A = \Omega + (\max_i \omega_i)I$ has nonnegative entries, Ω qualifies as an ML-matrix, named in honor of the mathematical economists Metzler and Leontief [41]. Typically Ω is irreducible in the sense that a particle of type i can ultimately generate descendents of type j for any pair i and j. If this is the case, then A is also irreducible, and the classical Perron–Frobenius theorem shows that A possesses an algebraically simple, positive eigenvalue γ that dominates the real part of any other eigenvalue [41]. Associated with γ are essentially unique left and right eigenvectors v and w having positive entries. It follows that Ω possesses the algebraically simple, real eigenvalue $\lambda = \gamma - \max_i \omega_i$ with v and w as left and right eigenvectors. Furthermore, any other eigenvalue θ of Ω satisfies $\mathrm{Re}(\theta) < \lambda$. The underlying branching process is said to be subcritical, critical, or supercritical according as λ is negative, zero, or positive, respectively.

In some applications, it is awkward to model reproduction as occurring simultaneously with death. Birth–death processes offer an attractive alternative. In a birth–death process, a type-i particle experiences death at rate μ_i and reproduction of offspring particles of type j at rate β_{ij}. Over its lifetime, a particle of type i in a birth–death process produces the vector $(n_1, \ldots, n_r)$ of offspring particles with

probability

$$\int_0^\infty \mu_i e^{-\mu_i t} \prod_{j=1}^r \frac{(\beta_{ij}t)^{n_j}}{n_j!} e^{-\beta_{ij}t} dt = \frac{\mu_i}{(\mu_i+\beta_i)^{\bar{n}+1}} \binom{\bar{n}}{n_1 \dots n_r} \prod_{j=1}^r \beta_{ij}^{n_j},$$

where $\beta_i = \sum_{j=1}^r \beta_{ij}$ and $\bar{n} = \sum_{i=1}^r n_i$. We can turn a birth–death process into a branching process in two ways. In the budding model, we view the parent particle as dying at each reproduction event and being replaced by an identical substitute. Reproduction continues until one of the substitutes dies an ordinary death. This interpretation of the birth–death process corresponds to a branching process with $\omega_i = \mu_i + \beta_i$ and progeny generating functions

$$P_i(s) = \frac{\mu_i}{\mu_i+\beta_i} + \frac{1}{\mu_i+\beta_i} \sum_{j=1}^r \beta_{ij} s_i s_j.$$

By contrast, in the bursting model, a parent particle collects its offspring and holds them for release at its death. In this approximation, $\omega_i = \mu_i$ and

$$P_i(s) = \sum_{k=0}^\infty \frac{\mu_i}{(\mu_i+\beta_i)^{k+1}} \left(\sum_{j=1}^r \beta_{ij} s_j \right)^k.$$

A brief calculation yields the progeny mean $f_{ij} = \frac{\partial}{\partial s_j} P_i(\mathbf{1}) = \beta_{ij}/\mu_i$ in the bursting model. Interpreting this ratio as a birth rate times an expected lifetime makes perfect sense. We can reverse our procedure and approximate a general branching process with progeny means f_{ij} and death rates ω_i by a birth–death process with birth rates $\beta_{ij} = \mu_i f_{ij}$ and death rates μ_i.

The preceding constructions work only when exponential lifetimes are assumed. In the non-Markovian Crump–Mode–Jagers branching process model, births can occur at multiple, arbitrary times during the parent's lifetime [10]. Kimmel and Axelrod [22] presented several biological examples of this more general branching process.

In many applications, immigration from external sources occurs in addition to birth and death. Two key assumptions simplify the mathematical treatment of immigration. The first states that immigration takes place independently of birth and death. The second states that the immigrants of each type enter according to independent Poisson processes. These strong assumptions are consistent with many applications. They do not require that immigration be homogeneous in time. In fact, we will feature several models where the rate of immigration grows or decays exponentially.

3. Poisson Processes and Campbell's Formulas. Poisson processes model the formation of random points in space or time [23]. Since immigration into a continuous-time branching process involves the arrival of particles in time, it would appear that we could confine our attention to the state space $[0, \infty)$. However, to accommodate marking, we will need to deal with more general state spaces [23]. Thus, the modern definition of Poisson processes involves a random set of points falling within some measurable space Φ equipped with a σ-finite intensity measure ν. If the set of random points is denoted by Π, and A is a measurable subset of Φ, then the counting random variable

$$N_A = \#\{\Pi \cap A\}$$

is either always finite or always countably infinite, depending on whether $\nu(A)$ is finite or infinite. To prevent random points from falling on top of one another, ν is required to possess no point masses (atoms). Equally important, any collection $\{N_{A_1}, \ldots, N_{A_k}\}$ of counting random variables is independent provided the corresponding subsets $A_1, \ldots, A_k$ are disjoint. From these simple assumptions, one can deduce that every N_A with $\nu(A) < \infty$ is Poisson distributed [23].

In most applications, the intensity measure $\nu(A) = \int_A \lambda(x)dx$ is determined by integrating an intensity function $\lambda(x)$ against ordinary length, area, or volume. In this setting, it is impossible for ν to contain a point mass. In practice, we interpret $\lambda(x)dx$ as the infinitesimal probability that a region around x of infinitesimal volume dx contains a random point. If $\lambda(x)$ is constant, then the corresponding process is said to be homogeneous. In Poisson processes constructed by marking, we encounter mixed continuous/discrete intensities that entail integration over some arguments and summation over other arguments.

Campbell's theorem deals with random sums of the sort

$$S = \sum_{X \in \Pi} f(X), \tag{1}$$

where the real-valued function $f : \Phi \to R$ is measurable and X is a generic point of Π [23]. One can show that the random sum (1) defining S converges absolutely with probability 1 if and only if

$$\int \min\{|f(x)|, 1\} d\nu(x) < \infty. \tag{2}$$

To evaluate the moments of S, consider the special case of a simple function $f(x) = \sum_{j=1}^m c_j 1_{A_j}(x)$ defined by a partition $A_1, \ldots, A_m$ of Φ. Because the sets A_j are disjoint, we can write

$$S = \sum_{j=1}^m c_j N_{A_j}.$$

This representation makes it clear that

$$\begin{aligned} \mathrm{E}\,(S) &= \sum_{j=1}^m c_j \,\mathrm{E}\,(N_{A_j}) \\ &= \sum_{j=1}^m c_j \int_{A_j} d\nu(x) \\ &= \int f(x) d\nu(x). \end{aligned} \tag{3}$$

If we use the independence of the N_{A_j} and the fact that $\mathrm{Var}(N_{A_j}) = \mathrm{E}\,(N_{A_j})$, then we deduce the similar formula

$$\begin{aligned} \mathrm{Var}(S) &= \sum_{j=1}^m c_j^2 \,\mathrm{E}\,(N_{A_j}) \\ &= \sum_{j=1}^m c_j^2 \int_{A_j} d\nu(x) \\ &= \int f(x)^2 d\nu(x). \end{aligned} \tag{4}$$

If we have a second sum $T = \sum_{X \in \Pi} g(X)$, then

$$\text{Cov}(S,T) = \int f(x)g(x)d\nu(x). \tag{5}$$

Finally, invoking the full Poisson distributions of the N_{A_j} leads to the formulas

$$\begin{aligned} \text{E}\,(e^{iuS}) &= \exp\Big\{ -\int [1 - e^{iuf(x)}]d\nu(x)\Big\}, \\ \text{E}\,(u^S) &= \exp\Big\{ -\int [1 - u^{f(x)}]d\nu(x)\Big\} \end{aligned} \tag{6}$$

for the characteristic function of S and for the probability generating function of S when $f(x)$ is nonnegative and integer valued. Appropriate limit arguments establish formulas (3) through (6) for more general functions $f(x)$.

In dealing with branching processes, we will exploit a device called *marking* for constructing one Poisson process from another [23]. In marking, we suppose that there is a second measurable space Γ of possible marks for each point $x \in \Phi$. For each random point $X = x \in \Pi$, a mark $y \in \Gamma$ is independently assigned by sampling from a distribution $p(x, \cdot)$ on Γ. If the distributions $p(x, \cdot)$ are compatible in the sense that $x \to p(x, B)$ is measurable for each measurable $B \subset \Gamma$, then we get a Poisson process on the product space $\Phi \times \Gamma$ with intensity measure

$$\gamma(C) = \int \int 1_C(x,y)p(x,dy)d\nu(x).$$

This marking theorem permits straightforward calculation of certain branching process moments and probabilities through the application of Campbell's formulas.

In a branching process with immigration, it is fruitful to model the immigrants of the different types as entering according to independent Poisson processes. If $\eta_i(t)$ denotes the intensity (or rate) of immigration of type-i particles, then we can view the overall immigration process as a marked Poisson process with intensity $\bar{\eta}(t) = \sum_{i=1}^{r} \eta_i(t)$. A random immigrant at time t is marked as a type-i particle with probability $\eta_i(t)/\bar{\eta}(t)$. This trivial kind of marking is sometimes called *coloring*. We will invoke a more complex form of marking that marks an immigrant at time τ with both its type N and its count of descendent particles $(Y_1, \ldots, Y_r)$ at some later time t.

4. Generating Functions for Multitype Processes. In a multitype, continuous-time, Markovian branching process, it is well known that the generating functions $Q_i(t,s)$ satisfy the backward system of ordinary differential equations

$$\frac{\partial}{\partial t} Q_i(t,s) = -\omega_i Q_i(t,s) + \omega_i P_i[Q(t,s)] \tag{7}$$

with initial conditions $Q_i(0,s) = s_i$. Although the nonlinear system (7) is usually impossible to solve analytically, it can be solved numerically and leads directly to the linear system of ordinary differential equations (15) for the mean vector $\text{E}\,(Z_t)$. Equation (15) is discussed later; for the sake of completeness, we now repeat the classic derivation of the backward system.

The argument proceeds via a preliminary system of integral equations. By assumption, the initial particle has random lifetime T_i with exponential distribution

$F_i(t) = 1 - e^{-\omega_i t}$. If we condition on the value of T_i, then it is clear that

$$Q_i(t,s) = \mathrm{E}\,(s^{Z_t} \mid T_i > t)[1 - F_i(t)] + \int_0^t \mathrm{E}\,(s^{Z_t} \mid T_i = \tau) dF_i(\tau). \tag{8}$$

On the one hand, if $T_i > t$, then the original particle is alive at time t and has no descendents. Hence,

$$\mathrm{E}\,(s^{Z_t} \mid T_i > t)[1 - F_i(t)] = s_i e^{-\omega_i t}.$$

On the other hand, if the original particle dies at time $\tau \leq t$, then it generates a random vector of offspring particles, each of which founds a separate clan that evolves independently of other clans during the remaining time $t - \tau$. By definition, $Q_k(t-\tau, s)$ is the generating function for the current descendents in a clan founded by a type-k offspring. These considerations imply that

$$\begin{aligned} \mathrm{E}\,(s^{Z_t} \mid T_i = \tau) &= \sum_j p_{ij} Q_1(t-\tau,s)^{j_1} \cdots Q_r(t-\tau,s)^{j_r} \\ &= P_i[Q(t-\tau,s)]. \end{aligned}$$

Substituting these results in (8) produces

$$Q_i(t,s) = s_i e^{-\omega_i t} + \int_0^t P_i[Q(t-\tau,s)]\omega_i e^{-\omega_i \tau} d\tau.$$

This integral equation can be simplified by multiplying by $e^{\omega_i t}$ and changing the variable of integration from τ to $t - \tau$. These steps yield the revised equation

$$Q_i(t,s)e^{\omega_i t} = s_i + \omega_i \int_0^t P_i[Q(\tau,s)]e^{\omega_i \tau} d\tau,$$

which, when differentiated with respect to t and multiplied by $e^{-\omega_i t}$, gives the backward equation (7).

In the presence of immigration, the backward system of equations (7) must be supplemented with additional differential equations. To derive the appropriate extension, assume constant Poisson immigration with intensity η_i per unit time for particles of type i. Let $R(t,s)$ denote the multivariate generating function for the total particles of different types starting from 0 particles at time 0. If $\bar{\eta} = \sum_i \eta_i$ is the overall immigration rate, then by conditioning on the arrival time τ of the first immigrant we can write the integral equation

$$R(t,s) = \int_0^t e^{-\bar{\eta}\tau} \sum_{i=1}^r \eta_i R(t-\tau,s) Q_i(t-\tau,s) d\tau. \tag{9}$$

Here the product $R(t-\tau,s)Q_i(t-\tau,s)$ summarizes the subsequent evolution of the process starting with a single type-i immigrant at time τ. If we multiply (9) by $e^{\bar{\eta}t}$, change the variable of integration from τ to $t - \tau$, differentiate with respect to t, multiply the result by $e^{-\bar{\eta}t}$, and rearrange, then we find that

$$\frac{\partial}{\partial t} R(t,s) = -\bar{\eta} R(t,s) + R(t,s) \sum_{i=1}^r \eta_i Q_i(t,s). \tag{10}$$

Equation (10) can be solved numerically in conjunction with the system (7). For a process starting with n_1 particles of type 1, n_2 particles of type 2, and so forth, the independent growth of all clans allows us to write the generating function

$$R(t,s)Q(t,s)^n = R(t,s)\prod_{i=1}^{r} Q_i(t,s)^{n_i}, \tag{11}$$

summarizing the total population at time t.

Even when the intensity of immigration is nonconstant, we can fall back on Campbell's theorem to derive an explicit expression for $R(t,s)$. Because $R(t,s)$ is the generating function of a vector-valued sum of the form (1), Campbell's formula (6) applies if we mark each immigrant particle by its type and numbers of descendents at time t. Given nonconstant immigration rates $\eta_i(t)$, it follows that

$$R(t,s) = e^{-\int_0^t \bar{\eta}(\tau)\sum_{i=1}^{r}\frac{\eta_i(\tau)}{\bar{\eta}(\tau)}[1-Q_i(t-\tau,s)]d\tau}. \tag{12}$$

To evaluate $R(t,s)$ when $\eta_i(t) = \eta_i e^{\beta_i t}$, consider the intermediate function

$$\begin{aligned} H_i(t,s) &= \int_0^t \eta_i(\tau)[1-Q_i(t-\tau,s)]d\tau \\ &= \eta_i e^{\beta_i t}\int_0^t e^{-\beta_i\tau}[1-Q_i(\tau,s)]d\tau, \end{aligned}$$

which can be differentiated to produce

$$\frac{\partial}{\partial t}H_i(t,s) = \beta_i H_i(t,s) + \eta_i[1-Q_i(t,s)] \tag{13}$$

with initial condition $H_i(0,s) = 0$. This equation and the system (7) can be simultaneously integrated numerically to give both the $Q_i(t,s)$ and the $H_i(t,s)$.

As a check on our calculations, we now recover (10) when immigration rates are constant. If we define $H(t,s) = \sum_i H_i(t,s)$ and set all $\beta_i = 0$, then summing equation (13) on i yields

$$\frac{\partial}{\partial t}H(t,s) = \sum_{i=1}^{r}\eta_i[1-Q_i(t,s)].$$

It follows that

$$\begin{aligned} \frac{\partial}{\partial t}R(t,s) &= -e^{-H(t,s)}\frac{\partial}{\partial t}H(t,s) \\ &= -e^{-H(t,s)}\sum_{i=1}^{r}\eta_i[1-Q_i(t,s)] \\ &= -\bar{\eta}R(t,s) + R(t,s)\sum_{i=1}^{r}\eta_i Q_i(t,s). \end{aligned}$$

5. Extinction. In the absence of immigration, $Q_i(t,\mathbf{0})$ is the probability that the process is extinct by time t starting from a single particle of type i at time 0. It is intuitively clear that $Q_i(t,\mathbf{0})$ monotonically increases to a limiting value x_i. The vector of extinction probabilities $x = (x_1,\ldots,x_r)$ can be characterized by setting

the right-hand side of (7) equal to 0. This gives the algebraic system $x = P(x)$ determining x. For both subcritical and critical processes, extinction is certain and $x = \mathbf{1}$. For a subcritical process, it is possible to prove this fact by constructing a Liapunov function $f(t)$ [19]. Let v be the left eigenvector corresponding to the dominant eigenvalue λ of Ω, and let $D(\omega)$ be the diagonal matrix with ith diagonal entry ω_i. Our Liapunov function is the inner product $f(t) = v[1 - Q(t,s)]^*$. Here and elsewhere, the superscript $*$ indicates a vector or matrix transpose. Equation (7), the mean value theorem, and the monotonicity of the differential $dP(s)$ together imply

$$\begin{aligned}
\frac{d}{dt} f(t) &= vD(\omega)\{Q(t,s) - P[Q(t,s)]\}^* \\
&= vD(\omega)[Q(t,s) - \mathbf{1}]^* + vD(\omega)\{P(\mathbf{1}) - P[Q(t,s)]\}^* \\
&\le vD(\omega)[Q(t,s) - \mathbf{1}]^* + vD(\omega)dP(\mathbf{1})[\mathbf{1} - Q(t,s)]^* \\
&= vD(\omega)[dP(\mathbf{1}) - I][\mathbf{1} - Q(t,s)]^* \\
&= v\Omega[\mathbf{1} - Q(t,s)]^* \\
&= \lambda v[\mathbf{1} - Q(t,s)]^*.
\end{aligned}$$

Because all entries of v are positive, $\lambda v[\mathbf{1} - Q(t,s)]^*$ is negative unless $Q(t,s) = \mathbf{1}$. It follows from Liapunov's theorem [17] that $Q(t,s)$ converges to $\mathbf{1}$, not just for $s = \mathbf{0}$ but for all s with $Q(t,s) \neq \mathbf{1}$. For a supercritical process, extinction is uncertain, and owing to irreducibility, all components x_i of x satisfy $x_i < 1$.

It is possible for extinction to occur in the presence of immigration if the rate of immigration falls rapidly enough. If we let t tend to ∞ in (12) and assume exponential immigration rates $\eta_i(t) = \eta_i e^{\beta_i t}$ with $\beta_i < 0$, then we find that

$$\begin{aligned}
R(\infty, \mathbf{0}) &= e^{-\int_0^\infty \sum_{i=1}^r \eta_i e^{\beta_i \tau}(1 - x_i) d\tau} \\
&= e^{\sum_{i=1}^r (1 - x_i)\eta_i/\beta_i}.
\end{aligned}$$

Starting from n_i type-i particles, $1 \le i \le r$, the independent behavior of initial clans and subsequent clans founded by immigrants entails an ultimate extinction probability of

$$\left(\prod_{i=1}^r x_i^{n_i}\right) e^{\sum_{i=1}^r (1 - x_i)\eta_i/\beta_i}.$$

6. Marginal Distributions. With these many results now in hand, we have all the ingredients for finding the coefficients of any marginal distribution of $R(t,s)$, $Q_i(t,s)$, or the product of these, such as that appearing in (11). The method is generic and depends on a simple application of Fourier analysis [16, 24]. Consider a nonnegative, integer-valued random variable N with probability generating $C(x) = \sum_{k=0}^\infty c_k x^k$. We can extract the coefficients $c_k = \Pr(N = k)$ by extending $C(x)$ to the boundary of the unit circle in the complex plane via the equation

$$C(e^{2\pi\sqrt{-1}y}) = \sum_{k=0}^\infty c_k e^{2\pi\sqrt{-1}ky}.$$

This creates a periodic function in y whose kth Fourier coefficient c_k can be recovered by the finite Riemann sum approximation

$$c_k \approx \frac{1}{n} \sum_{j=0}^{n-1} C(e^{2\pi\sqrt{-1}j/n}) e^{-2\pi\sqrt{-1}kj/n} \tag{14}$$

for some large power n of 2. In practice, one evaluates this finite Fourier transform using the fast Fourier transform algorithm [37]. For a sufficiently large power n, all of the coefficients $c_0, \ldots, c_{n-1}$ can be computed accurately. Accuracy can be checked by comparing the numerically computed mean and variance of N with its theoretical mean and variance when these are available.

In the current setting, if we seek the kth marginal distribution of, say $R(t,s)$, then we set all $s_j = 1$ except for s_k, which we vary over the boundary of the unit circle. We then numerically integrate the differential equations for the generating functions $Q_i(t,s)$ and $H_i(t,s)$ and apply (12) for each boundary value of s_k. Finally, we take the finite Fourier transform to recover the coefficients of the marginal distribution. The method works best if the marginal distribution is fairly concentrated around 0. This is the case for most subcritical processes or subcritical processes renewed by moderate immigration. It fails for supercritical processes if the time index t is sufficiently large for explosive growth to have taken hold.

7. Calculation of Means. To recover the expected vector $\mathrm{E}(Z_t)$, one can differentiate the backward equation (7) with respect to s_i and set $s = \mathbf{1}$. We can arrive at the same result and better accommodate immigration by deriving forward equations. Let $m(t) = [m_1(t), \ldots, m_r(t)]$ be the vector of expected particle counts at time t. As before, let $f_{ij} = \frac{\partial}{\partial s_j} P_i(\mathbf{1})$ be the expected number of offspring particles of type j born to a particle of type i. Since particles of type i at time $t+\tau$ arise from (a) particles of type i at time t that do not die during $(t, t+\tau)$, (b) particles of type j that die during $(t, t+\tau)$ and reproduce particles of type i, or (c) immigration of type-i particles, we find that

$$m_i(t+\tau) = m_i(t)(1-\omega_i\tau) + \sum_{j=1}^{r} m_j(t)\omega_j f_{ji}\tau + \eta_i(t)\tau + o(\tau),$$

where $\eta_i(t)$ is the immigration rate of type-i particles at time t. Forming the corresponding difference quotients and sending τ to 0 yield the forward system of differential equations

$$\frac{d}{dt} m_i(t) = \sum_{j=1}^{r} m_j(t)\omega_j(f_{ji} - 1_{\{j=i\}}) + \eta_i(t), \tag{15}$$

which we summarize in vector-matrix notation as

$$\frac{d}{dt} m(t) = m(t)\Omega + \eta(t),$$

employing the immigration vector $\eta(t) = [\eta_1(t), \ldots, \eta_r(t)]$ and the $r \times r$ matrix $\Omega = [\omega_i(f_{ij} - 1_{\{i=j\}})]$. It is straightforward to check that this ordinary differential equation has solution

$$m(t) = m(0)e^{t\Omega} + \int_0^t \eta(\tau)e^{(t-\tau)\Omega} d\tau \tag{16}$$

involving the matrix exponential function. The convolution integral appearing in formula (16) can be evaluated when the rates of immigration $\eta_i(t) = \alpha_i e^{\beta_i t}$ are exponential. In this case, we exploit the representation

$$\eta(t) \;=\; \mathbf{1}\sum_{i=1}^{r} \alpha_i u_i^* u_i e^{\beta_i t} \;=\; \mathbf{1}\sum_{i=1}^{r} \alpha_i u_i^* u_i e^{\beta_i t I}$$

of the row vector $\eta(t)$ using the unit vectors u_i and their transposes. It follows that

$$\begin{aligned}\int_0^t \eta(\tau)e^{(t-\tau)\Omega}d\tau &= \mathbf{1}\sum_{i=1}^r \alpha_i u_i^* u_i \int_0^t e^{\beta_i \tau I}e^{(t-\tau)\Omega}d\tau \\ &= \mathbf{1}\sum_{i=1}^r \alpha_i u_i^* u_i \int_0^t e^{\tau(\beta_i I-\Omega)}d\tau e^{t\Omega} \\ &= \mathbf{1}\sum_{i=1}^r \alpha_i u_i^* u_i (\Omega-\beta_i I)^{-1}\left(e^{t\Omega}-e^{t\beta_i I}\right).\end{aligned}$$

We can safely substitute the result in (16) provided no β_i is an eigenvalue of Ω. One can generalize these formulas to other rate functions. For example, if $\eta_i(t) = \alpha_i \cos(\beta_i t) + \gamma_i$ with $|\alpha_i| \leq \gamma_i$, then the convolution integral can be evaluated by substituting $\cos(\beta_i t) = (e^{\sqrt{-1}\beta_i t} + e^{-\sqrt{-1}\beta_i t})/2$ and proceeding as before. The choice $\eta_i(t) = \alpha_i t e^{\beta_i t}$ can be handled via integration by parts.

For a subcritical process with constant immigration rates, a stochastic equilibrium is ultimately reached between extinction and immigration. The mean $-\alpha\Omega^{-1}$ of the equilibrium distribution is obtained by taking limits in (16) or by setting the left-hand side of (15) equal to 0 and solving.

8. The Vec Operator and Kronecker Products. To simplify our derivation of variances in the next section, we will use the Vec operator and matrix Kronecker products. The Vec operator stacks the successive columns of a matrix to form a column vector. If $A = (a_{ij})$ is a $k \times l$ matrix and $B = (b_{ij})$ is an $m \times n$ matrix, then the Kronecker product $A \otimes B$ is the $km \times ln$ block matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1l}B \\ \vdots & \ddots & \vdots \\ a_{k1}B & \cdots & a_{kl}B \end{pmatrix}.$$

The Kronecker product, which also goes by the names tensor product and direct product, should be distinguished from the direct sum. The direct sum of the matrices A and B is a block diagonal matrix with A and B as diagonal blocks. The references [18, 29] cover many theoretical properties of Kronecker products.

The basic connection between the Vec operator and the Kronecker product is supplied by the identity

$$\operatorname{Vec}(ABC) = (C^* \otimes A)\operatorname{Vec}(B) \tag{17}$$

for compatible matrices A, B, and C. Two other conversion formulas are also worth mentioning. For two finite sequences of column vectors v_i and w_i related by $w_i = Av_i$, the column vectors w and v produced by stacking satisfy

$$w = (I \otimes A)v \tag{18}$$

for an identity matrix I of the correct dimension. Similarly, if $w_i = \sum_j a_{ij}v_j$ for column vectors v_j and w_i, then stacking yields

$$w = (A \otimes I)v. \tag{19}$$

9. Calculation of Variances. The variance matrix $V_i(t) = \mathrm{E}\,(Z_t^* Z_t) - \mathrm{E}\,(Z_t)^*\,\mathrm{E}\,(Z_t)$ of the particle count vector Z_t can also be found by deriving and solving a forward differential equation [2, 25]. Here $V_i(t)$ is subscripted by i to remind us that the process starts from a single type-i particle. To avoid breaking the flow of our current discussion, we defer several lengthy proofs to the appendix. Arguments there show that if we stack the vectors $\mathrm{Vec}[V_i(t)]$, then the resulting vector $\mathrm{Vec}[V(t)]$ reduces to the integral

$$\mathrm{Vec}[V(t)] = \int_0^t e^{\tau\Omega} \otimes e^{(t-\tau)\Omega^*} \otimes e^{(t-\tau)\Omega^*} d\tau\, \mathrm{Vec}(C), \tag{20}$$

where the matrix C is given by (A.4) in the appendix. When Ω is diagonalizable, the integral (20) and its counterpart (A.3) in the appendix can be evaluated via (A.5).

In the presence of immigration, evaluation of $\mathrm{Var}(Z_t)$ proceeds via Campbell's formulas (4) and (5). For the sake of simplicity, temporarily assume that we start with no particles and only type-i particles can immigrate. Let T denote a Poisson time of immigration and Y_T the vector of descendents of such an immigrant at time t. In this setting, Z_t reduces to the random sum

$$Z_t = \sum_{(T,Y_T)\in\Pi} Y_T$$

over the points (T, Y_T) of a marked Poisson process Π. According to formulas (4) and (5),

$$W_i(t) = \mathrm{Var}(Z_t) = \int_0^t \eta_i(\tau)\,\mathrm{E}\,(Y_T^* Y_T \mid T = \tau) d\tau.$$

Since $\mathrm{E}\,(Y_T \mid T = \tau) = u_i e^{(t-\tau)\Omega}$ and

$$\mathrm{E}\,(Y_T^* Y_T \mid T = \tau) = V_i(t-\tau) + e^{(t-\tau)\Omega^*} u_i^* u_i e^{(t-\tau)\Omega},$$

it follows that

$$W_i(t) = \int_0^t \eta_i(\tau) \left[V_i(t-\tau) + e^{(t-\tau)\Omega^*} u_i^* u_i e^{(t-\tau)\Omega} \right] d\tau.$$

Stacking the columns of $W_i(t)$ consequently produces

$$\mathrm{Vec}[W_i(t)] = \int_0^t \eta_i(\tau) \left\{ \mathrm{Vec}[V_i(t-\tau)] + e^{(t-\tau)\Omega^*} \otimes e^{(t-\tau)\Omega^*}\, \mathrm{Vec}(u_i^* u_i) \right\} d\tau.$$

When Ω is diagonalizable in the form $\Omega = A\Delta A^{-1}$ and $\eta_i(t) = \alpha_i e^{\beta_i t}$, the matrix integral

$$\int_0^t \eta_i(\tau) e^{(t-\tau)\Omega^*} \otimes e^{(t-\tau)\Omega^*} d\tau$$

is similar via $(A^*)^{-1} \otimes (A^*)^{-1}$ to the diagonal matrix with diagonal entry at level (j,k) of

$$\alpha_i \int_0^t e^{\beta_i \tau} e^{(t-\tau)\delta_j} e^{(t-\tau)\delta_k} d\tau = \alpha_i \frac{e^{\beta_i t} - e^{(\delta_j+\delta_k)t}}{\beta_i - \delta_j - \delta_k},$$

where δ_j is the jth diagonal entry of Δ. To calculate the vector integral

$$\int_0^t \eta_i(\tau)\,\mathrm{Vec}[V_i(t-\tau)]d\tau = \int_0^t \eta_i(t-\tau)\,\mathrm{Vec}[V_i(\tau)]d\tau,$$

we employ (20) and the representation

$$\mathrm{Vec}[V_i(\tau)] = u_i \otimes I \otimes I\,\mathrm{Vec}[V(\tau)].$$

It therefore suffices to evaluate

$$\begin{aligned}
&\int_0^t \eta_i(t-\tau)\,\mathrm{Vec}[V(\tau)]d\tau \\
&= \int_0^t \eta_i(t-\tau) \int_0^\tau e^{\iota\Omega} \otimes e^{(\tau-\iota)\Omega^*} \otimes e^{(\tau-\iota)\Omega^*} d\iota d\tau\,\mathrm{Vec}(C) \\
&= \alpha_i \int_0^t \int_0^\tau e^{(t-\tau)\beta_i I+\iota\Omega} \otimes e^{(\tau-\iota)\Omega^*} \otimes e^{(\tau-\iota)\Omega^*} d\iota d\tau\,\mathrm{Vec}(C).
\end{aligned}$$

This matrix double integral is similar via $H = A \otimes (A^*)^{-1} \otimes (A^*)^{-1}$ to a diagonal matrix double integral whose diagonal entry at level (i,j,k) turns out to be

$$\frac{1}{\delta_i - \delta_j - \delta_k}\left(\frac{e^{\delta_i t} - e^{\beta_i t}}{\delta_i - \beta_i} - \frac{e^{(\delta_j+\delta_k)t} - e^{\beta_i t}}{\delta_j + \delta_k - \beta_i}\right). \tag{21}$$

For a subcritical process with constant immigration, the diagonal entry (21) tends to $\delta_i^{-1}(\delta_j + \delta_k)^{-1}$ as t tends to ∞ and permits explicit evaluation of the variance of the equilibrium distribution. Finally, when the process starts with count vector $(n_1, \ldots, n_r)$ and immigration simultaneously feeds into each of the types, independent evolution of the various clans gives an overall variance matrix of

$$\mathrm{Var}(Z_t) = \sum_{i=1}^r n_i V_i(t) + \sum_{i=1}^r W_i(t).$$

10. Total Descendents. There are many other interesting random vectors in addition to the vector Z_t of particle counts. For example, in a subcritical process it makes sense to consider the ultimate number of particles Y_{ij} of type j attributable to an ancestral particle of type i. The original particle is included in this accounting when $j = i$. If $a_{ij} = \mathrm{E}\,(Y_{ij})$, then the recurrence

$$a_{ij} = 1_{\{i=j\}} + \sum_{k=1}^r f_{ik} a_{kj} \tag{22}$$

follows by conditioning on the reproductive outcome of the ancestor. If we collect the a_{ij} into a matrix A, then the matrix version of (22) is $A = I + FA$ with solution

$$A = (I - F)^{-1}. \tag{23}$$

To calculate variances, let Y_i be the vector $(Y_{i1}, \ldots, Y_{ir})$ and X the count vector $(X_1, \ldots, X_r)$ for the offspring of a type-i ancestor. Conditioning on X yields

$$B_i = \mathrm{Var}(Y_i) = \mathrm{Var}[\mathrm{E}\,(Y_i \mid X)] + \mathrm{E}\,[\mathrm{Var}(Y_i \mid X)].$$

Because each offspring particle possesses the same distribution of descendents as the ancestor, we have

$$\begin{aligned} B_i &= \operatorname{Var}\left[X(I-F)^{-1}\right] + \mathrm{E}\left(\sum_{j=1}^{r} X_j B_j\right) \\ &= (I-F^*)^{-1} G_i (I-F)^{-1} + \sum_{j=1}^{r} f_{ij} B_j, \end{aligned}$$

where G_i is the variance of X. Stacking the columns of the matrices on both sides of this equation gives

$$\operatorname{Vec}(B_i) = (I-F^*)^{-1} \otimes (I-F^*)^{-1} \operatorname{Vec}(G_i) + \sum_{j=1}^{r} f_{ij} \operatorname{Vec}(B_j),$$

and stacking the resulting vectors $\operatorname{Vec}(B_i)$ in turn gives

$$\operatorname{Vec}(B) = I \otimes (I-F^*)^{-1} \otimes (I-F^*)^{-1} \operatorname{Vec}(G) + F \otimes I \otimes I \operatorname{Vec}(B).$$

This last equation has solution

$$\operatorname{Vec}(B) = (I-F)^{-1} \otimes (I-F^*)^{-1} \otimes (I-F^*)^{-1} \operatorname{Vec}(G). \tag{24}$$

11. Sensitivity. In many practical problems, it is desirable to know how sensitive model predictions are to changes in various parameters. Such information can have implications for public policies such as large-scale vaccination. In many cases, underlying model parameters cannot be measured directly. Local sensitivity is measured by the partial derivatives of summary indices such as the extinction vector x and the dominant eigenvalue λ. Both of these indices are functions of the underlying kinetic parameters of the branching process.

To determine the local sensitivity of λ to some parameter θ, suppose its left and right eigenvectors v and w are normalized so that $vw = 1$. Differentiating the identity $\Omega w = \lambda w$ with respect to θ yields

$$\left(\frac{\partial}{\partial\theta}\Omega\right) w + \Omega \frac{\partial}{\partial\theta} w = \left(\frac{\partial}{\partial\theta}\lambda\right) w + \lambda \frac{\partial}{\partial\theta} w.$$

If we multiply this by v on the left and invoke the identities $v\Omega = \lambda v$ and $vw = 1$, then we find that

$$\frac{\partial}{\partial\theta}\lambda = v\left(\frac{\partial}{\partial\theta}\Omega\right) w. \tag{25}$$

The sensitivity of v and w can be determined by an extension of this reasoning [29].

To find the partial derivative of the extinction vector x with respect to θ, we assume that the branching process is supercritical and resort to implicit differentiation of the equation $x(\theta) = P[x(\theta), \theta]$ derived by setting the left-hand side of the backward equation (7) equal to 0. The chain rule gives

$$\frac{\partial}{\partial\theta} x(\theta) = \frac{\partial}{\partial x} P[x(\theta), \theta] \frac{\partial}{\partial\theta} x(\theta) + \frac{\partial}{\partial\theta} P[x(\theta), \theta].$$

This equation has the obvious solution

$$\frac{\partial}{\partial\theta}x(\theta) = \left\{I - \frac{\partial}{\partial x}P[x(\theta),\theta]\right\}^{-1}\frac{\partial}{\partial\theta}P[x(\theta),\theta].$$

The indicated inverse does in fact exist, but the proof presents too much of a detour for our current purposes.

Finally, we can determine the local sensitivity of the expected numbers of total descendents by differentiating the equation $A = (I-F)^{-1}$. The result,

$$\frac{\partial}{\partial\theta}A = (I-F)^{-1}\frac{\partial F}{\partial\theta}(I-F)^{-1}, \tag{26}$$

depends on the sensitivity of the expected offspring matrix F.

12. Examples. We now turn to four sample applications of the methods reviewed. We begin with a simple paper-and-pencil example. Our subsequent examples require the help of a computer and standard mathematical software. Our first example models the transmission of an X-linked recessive disease and reports the mean and variance of the total number of descendents of a single carrier female or a single affected male as derived from (23) and (24). Our second example investigates the distribution of haplotypes in a growing human population and demonstrates via (14) the calculation of marginal distributions for a branching process with Poisson immigration of exponentially increasing intensity. Our third example employs an epidemic model to examine the local sensitivity of the total outbreak size to model parameters through (26). Finally, our fourth example uses data on the fitness of HIV single, double, and triple viral mutants to predict the mean and variance of the number of mutant viruses as a function of time, illustrating (16) and (20).

12.1. X-Linked Recessive Diseases. A rare X-linked recessive disease such as Becker's muscular dystrophy can be modeled as a 2-type branching process with type-1 particles being carrier females and type-2 particles being affected males [12, 13, 27]. For the sake of simplicity, we assume that carrier females average 2 children, and affected males 2θ children, where $\theta < 1$ specifies the fitness reduction experienced by affected males. In Becker's muscular dystrophy, $\theta \approx 1/2$. Because both types of individuals can produce normal children who do not carry the gene, we must carefully distinguish the generating functions O_1 and O_2 for the total number of children and the progeny generating functions

$$P_1(s_1,s_2) = O_1\Big(\frac{1}{2} + \frac{1}{4}s_1 + \frac{1}{4}s_2\Big),$$
$$P_2(s_1,s_2) = O_2\Big(\frac{1}{2} + \frac{1}{2}s_1\Big)$$

of the branching process. Here we suppose that children are equally divided between boys and girls. Affected males possess an X and a Y chromosome. They must pass the Y chromosome to their sons, so no sons are affected. They always pass the mutation-bearing X chromosome to their daughters, so all of the daughters are carriers. Carrier females pass the mutation-bearing chromosome with probability 1/2 to each child, regardless of its sex.

It simplifies matters to assume that O_1 and O_2 are Poisson generating functions. In this case, the numbers of carrier daughters and affected sons born to either type

parent are independent and Poisson distributed. Furthermore, the progeny mean matrix is

$$F = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \theta & 0 \end{pmatrix},$$

and the progeny variance matrices are

$$G_1 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}, \qquad G_2 = \begin{pmatrix} \theta & 0 \\ 0 & 0 \end{pmatrix}.$$

In this subcritical branching process, one can straightforwardly calculate the matrix

$$(I - F)^{-1} = \frac{1}{1-\theta} \begin{pmatrix} 2 & 1 \\ 2\theta & 1 \end{pmatrix}$$

determining the mean number of descendents—including children, grandchildren, and so forth—who inherit a mutant gene from a single ancestral carrier female or affected male. The matrices specifying the variances and covariances in these numbers can also be found with considerably more work. Based on (24), these turn out to be

$$B_1 = \frac{1}{(1-\theta)^3} \begin{pmatrix} 4[1+\theta+\theta^2] & 2+4\theta \\ 2+4\theta & 2+\theta \end{pmatrix}$$

and

$$B_2 = \frac{1}{(1-\theta)^3} \begin{pmatrix} 8\theta+4\theta^3 & 4\theta+2\theta^2 \\ 4\theta+2\theta^2 & 3\theta \end{pmatrix}.$$

12.2. A Haplotype Model. Fan and Lange [8, 26] consider a 4-type model of haplotype evolution for an autosomal dominant disease. This model sheds light on the merits of haplotype mapping versus linkage mapping. The primary distinction between these two forms of disease-gene mapping is that haplotype mapping exploits old, unobserved recombination events, while linkage mapping exploits contemporary recombination events.

The haplotype branching process commences with a new mutation at the disease locus. This mutation is carried on a chromosome background that changes over time as the chromosome suffers genetic recombination in passing from affected parents to affected children. Figure 1 depicts four different types of chromosomes bearing the mutation. The origin of the figure represents the genetic map position of the disease locus. To follow the disruption of the original mutant chromosome by recombination, we arbitrarily choose points to the left and right at map positions $-a$ and b. In type-1 individuals, the original mutant chromosome is intact over the whole interval $[-a, b]$. The mutant chromosomes of type-2 and type-3 individuals are disrupted by recombination on $[0, b]$ and $[-a, 0]$, respectively. The mutant chromosomes of type-4 individuals are disrupted on both intervals.

Affected individuals have reduced fitness compared to ordinary people. This fitness effect is reflected in the smaller mean of their common offspring generating function O. Under Haldane's model of recombination, recombination events occur on the mutant chromosome according to a Poisson process. Thus, the segments $[0, b]$ and $[-a, 0]$ are passed intact with probabilities $\hat{b} = e^{-b}$ and $\hat{a} = e^{-a}$, respectively. These

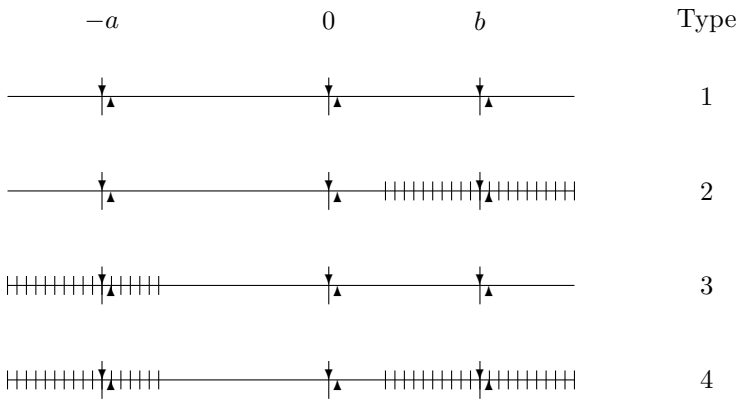

Fig. 1 *Four-type mutation model.*

considerations produce the progeny generating functions

$$P_1(s) = O\left(\frac{1}{2} + \frac{1}{2}\hat{a}\hat{b}s_1 + \frac{1}{2}\hat{a}[1-\hat{b}]s_2 + \frac{1}{2}[1-\hat{a}]\hat{b}s_3 + \frac{1}{2}[1-\hat{a}][1-\hat{b}]s_4\right),$$
$$P_2(s) = O\left(\frac{1}{2} + \frac{1}{2}\hat{a}s_2 + \frac{1}{2}[1-\hat{a}]s_4\right),$$
$$P_3(s) = O\left(\frac{1}{2} + \frac{1}{2}\hat{b}s_3 + \frac{1}{2}[1-\hat{b}]s_4\right),$$
$$P_4(s) = O\left(\frac{1}{2} + \frac{1}{2}s_4\right).$$

If reproduction is to take place simultaneously with death, then the common life expectancy ω^{-1} of people bearing the various chromosome types should be equated with the average age of parenthood, say 25 years.

In actual human populations, new mutants constantly refresh this process and constitute a form of immigration. Although we can safely assume that the stream of new immigrants is Poisson, the intensity of immigration depends directly on the size of the surrounding normal population, which is subject to growth and decline. Let us suppose, for the sake of simplicity, exponential growth at rate β. Only immigrants of type 1 occur. The constant α in the overall rate of immigration $\alpha e^{\beta t}$ is determined by the mutation rate μ, the size m_0 of the original normal population, and the average number of children c of a normal person according to the formula $\alpha = 2\mu\frac{1}{2}\omega m_0 c$. Here the factor 2 arises because each child inherits two homologous chromosomes, one maternal and one paternal in origin, either of which can mutate. The factor of $1/2$ arises because we double count children in attributing them to both their fathers and mothers, implicitly assuming a sex ratio of 1:1. The constants β and c are related by the identity $\beta = (\frac{c}{2}-1)\omega$ since reproduction must be adjusted for death to compute the net rate of exponential growth.

This model has been thoroughly analyzed in [8, 26] except for computation of the marginal distributions of the different types. Here we would like to focus on this numerical exercise using the demographic parameters of Finland, an interesting genetic isolate [15]. The current approximately 5 million Finns trace back to a founding

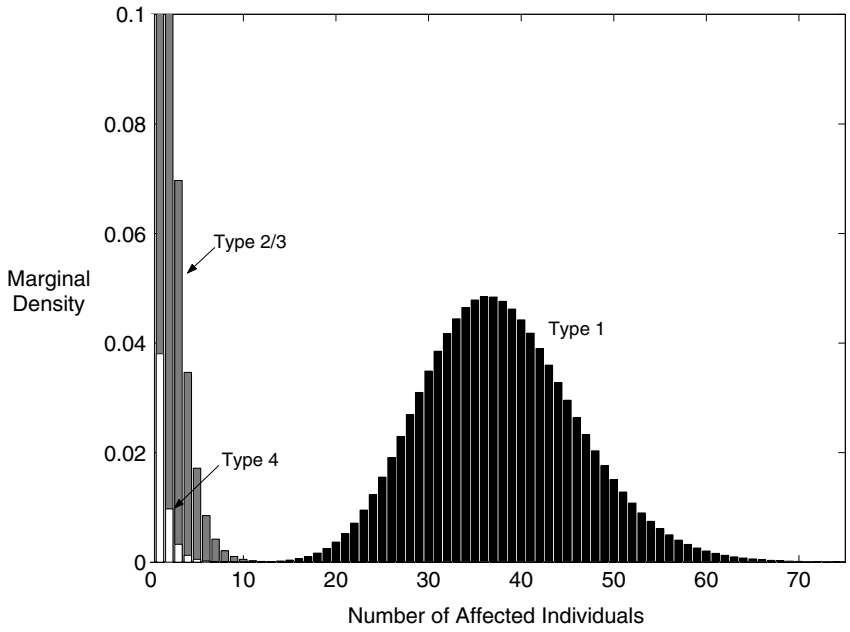

Fig. 2 *Marginal densities for the haplotype model.*

population of perhaps $m_0 = 1000$ people 2000 years ago. If we assume steady exponential growth, then these figures imply an exponential growth rate of $\beta = 4.3 \times 10^{-3}$ per year. Dominant diseases with high fitness are the most amenable to gene mapping by haplotype disruption, so we assume a fitness of $\theta = 0.75$ and a mutation rate of $\mu = 1.0 \times 10^{-6}$. Given these parameters and a generation time of $\omega^{-1} = 25$ years, we can easily compute the average number of offspring per normal individual $c = 2.2$ and the immigration constant $\alpha = 3.5 \times 10^{-5}$. For distances $a = b = 0.01$, Figure 2 plots the distribution of the four types at the present time $t = 2000$. These distributions have means $\mathrm{E}(Z_{2000,1}) = 37.74$, $\mathrm{E}(Z_{2000,2}) = \mathrm{E}(Z_{2000,3}) = 1.09$, and $\mathrm{E}(Z_{2000,4}) = 0.08$, for an average total of 40 affected individuals [8, 26]. The corresponding variances are $\mathrm{Var}(Z_{2000,1}) = 69.07$, $\mathrm{Var}(Z_{2000,2}) = \mathrm{Var}(Z_{2000,3}) = 2.09$, and $\mathrm{Var}(Z_{2000,4}) = 0.15$ when the offspring generating function O arises from a birth–death process.

12.3. Vaccination Strategies. Stochastic models of infectious disease and epidemics have been pursued for many decades. Both empirical studies and sophisticated models suggest that infectious diseases spread more easily between members of the same household than they do between members of a community at large [5]. This fact can dramatically impact the success of a vaccination program. To explore various vaccination strategies, we now discuss a branching process model whose states are households of infective and susceptible individuals from an extended community [4, 6]. The following are the premises of the model:

1. The process begins with a single primary infective person in a single household.
2. An infection involves no latency period and persists an exponential length of time in any one individual, who either ultimately dies or is cured and rendered immune to the disease.

3. Infections spread by pairwise encounters between infectives and susceptibles.
4. Infections spread more easily between members of the same household than between members of the community at large.
5. Households are selected randomly for vaccination, and some members of each selected household are randomly vaccinated. Vaccination confers complete protection.

These assumptions must be fleshed out with parameters. Let γ^{-1} be the mean duration of the infective period, β_H the infection rate per infective-susceptible pair within a household, β_C the infection rate per infective-susceptible pair within the broader community, ν_H the proportion of households chosen for vaccination, and ν_I the proportion of individuals vaccinated within chosen households. Among the households of the community, h_k contain exactly k members, giving $\mu_H = \sum_{k=1}^{m} kh_k$ community members in all. Here m is the maximum number of members in any household.

The states of the branching process are households summarized by ordered pairs (i, j), with i indicating the number of infective members of the household and j the number of susceptible members. In this accounting, we omit household members who have been vaccinated, who have recovered from infection, or who have died from infection. In view of our assumptions, $i+j \leq m$, $i \geq 1$, and $j \geq 0$. There are four kinds of events. (a) Households with a single infective, represented by $(1, j)$, drop out at rate γ because the infective is cured or dies. (b) This case is to be distinguished from households of type (i, j), with $i > 1$, that lose an infective at rate $i\gamma$ but thereafter remain in play as households of type $(i-1, j)$. (c) Households gain a new infective by secondary infection at rate $ij\beta_H$. This action converts a household of type (i, j) into a household of type $(i+1, j-1)$, assuming $j > 0$. (d) Finally, there is the possibility of a household member infecting a person outside the household. This converts a type-(i, j) household into a type-(i, j) household and a type-$(1, k-1)$ household, assuming the new infective resided in a household with k susceptibles. This branching process describes either the early stages of an epidemic when effectively all households are uninfected or all stages of a small outbreak touching only a fraction of community households.

Events of type (d) occur at rate $ik\beta_C\theta_k$, where θ_k depends on the fraction of households and people within households who are vaccinated. Because the number of people vaccinated within a household is binomially distributed with success probability ν_I, we have

$$\theta_k = (1-\nu_H)\frac{h_k}{\mu_H} + \nu_H \sum_{j=k}^{m} \frac{h_j}{\mu_H} \binom{j}{k} (1-\nu_I)^k \nu_I^{j-k}.$$

Summarizing our discussion so far, in state (i, j) the death rate is

$$\omega_{ij} = i\gamma + ij\beta_H + \sum_{k=1}^{m} ik\beta_C\theta_k,$$

and the progeny generating function is

$$P_{ij}(s) = \frac{i}{\omega_{ij}} \left(\gamma 1_{\{i=1\}} + \gamma 1_{\{i>1\}} s_{i-1,j} + j\beta_H s_{i+1,j-1} + \beta_C \sum_{k=1}^{m} k\theta_k s_{ij} s_{1,k-1} \right).$$

For purposes of illustration, we consider a novel application of the model to classical data on the spread of a respiratory disease on the island of Tristan de Cunha

Table 1 *Parameters for a household vaccination model.*

Parameter	Meaning	Value
β_H	Infection rate within households	0.021 days^{-1}
β_C	Infection rate between households	0.00056 days^{-1}
γ^{-1}	Average infectious period	5.5 days
m	Maximum household size	8
h_1	Households with 1 member	5
h_2	Households with 2 members	9
h_3	Households with 3 members	24
h_4	Households with 4 members	16
h_5	Households with 5 members	8
h_6	Households with 6 members	3
h_7	Households with 7 members	2
h_8	Households with 8 members	3

between 1964 and 1967 [6, 7]. Previous work has demonstrated a difference in infectivity rates between households versus within households [7]. Table 1 lists a between-household infection rate of 0.00056, a within-household infection rate of 0.021, and a mean duration of an infection of 5.5 days. We ignore differences in infectivity by age and the short latent period of approximately 1 day. Household sizes range from 1 to 8 persons.

For these parameter choices, the Tristan de Cunha epidemic is subcritical and ultimately self-contained. Nevertheless, a well-conceived vaccination strategy could reduce the total number of people affected by the disease during an outbreak. Based on (26) and assuming random selection of the primary infective, we can calculate the local sensitivity of the total outbreak size to changes in the parameters v_H and v_I. Figure 3 plots the gradient of total outbreak size with respect to v_H and v_I. In the region to the left of the dividing curve, the ultimate size of an outbreak is more sensitive to v_I; in the region to the right of the dividing curve, the ultimate size of an outbreak is more sensitive to v_H. These regions are not symmetric. In particular, when vaccination coverage v_I within households is good, say above 75%, it is always more effective to increase the proportion v_H of households covered than it is to increase v_I. These conclusions could be sharpened by taking into account the relative costs of acquiring new households for vaccination versus acquiring new members within participating households. Presumably, one could even push this kind of cost/benefit analysis to the point of choosing a single best combination of v_H and v_I.

12.4. HIV Resistance. The human immunodeficiency virus type 1 (HIV-1) that infects cells of the immune system is highly prone to mutate when its native RNA core is transcribed into DNA. Hence, over time, infected people harbor many different viral variants. The HIV literature distinguishes wild-type variants that evolve in the absence of antiviral drugs from mutant variants that evolve under the selective pressure of antiviral drugs. Several mutations make the virus more fit in the presence of particular antiviral drugs but less fit in the absence of the same drugs. For example, with the antiviral drug ritonavir that targets the viral protease gene, nearly complete resistance is achieved with three mutations, alanine at amino acid 82 (82A), valine at amino acid 71 (71V), and valine at amino acid 54 (54V) [33]. The possible mutants and the mutation events connecting them are shown in Figure 4. Molla et al. [33]

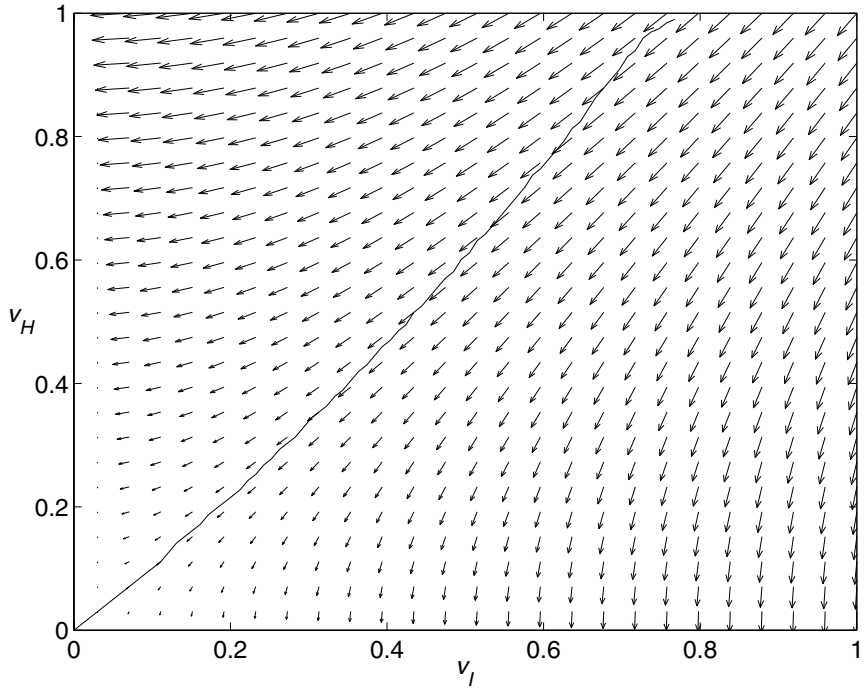

Fig. 3 *Gradient plot of total epidemic size.*

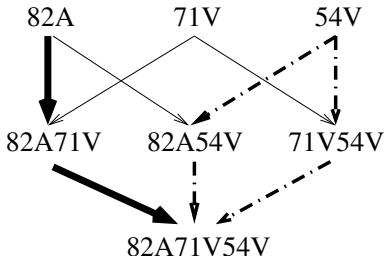

Fig. 4 *HIV resistance mutations for the antiviral ritonavir.*

found ritonavir-treated patients tend to accumulate mutants in a particular order, first 82A, then 71V, and finally 54V. The solid, dark arrows in Figure 4 trace this dominant pathway to resistance.

Deterministic mathematical models have been frequently applied to problems in HIV research, including the emergence of resistant variants [35]. More recent resistance models have also considered stochastic effects [36, 40]. We find it instructive to model viral infection and mutation as a branching process with six particle types, three types of viruses and three types of infected CD4 cells. Particle types 1, 2, and 3 are viruses with the accumulated mutations 82A, 82A71V, and 82A71V54V, respectively. Wild-type viruses are relatively rare under treatment conditions and are ignored in our model. Mutation occurs at rate m during transcription of the genome

Table 2 *Particle types in the HIV model.*

Type number	Virus or cell	Accumulated mutations	Death rate	Initial number
1	Virus	82A	v	8×10^3
2	Virus	82A71V	v	6
3	Virus	82A71V54V	v	< 1
4	Cell	82A	$c + f_4bc$	2×10^3
5	Cell	82A71V	$c + f_5bc$	22
6	Cell	82A71V54V	$c + bc$	< 1

of the infecting virus. It is appropriate to label an infected cell by the type of its transcribed viral genome rather than by the type of its infecting virus. Thus, CD4 cell types 4, 5, and 6 harbor viral genomes of types 1, 2, and 3. Table 2 summarizes the six types.

Viruses vanish at rate v and infected cells die at rate c. Viruses have a chance u of being cleared by the immune system before infecting a cell. During antiviral therapy, cells bud off new viruses at rates f_4bc, f_5bc, and bc, where the fitnesses $0 < f_4 < f_5 < 1$ scale the production of viruses for the different cell types. On average, the three cell types ultimately produce f_4b, f_5b, and b viruses. In the branching process interpretation of budding, the three cell types "die" at rates $c + f_4bc$, $c + f_5bc$, and $c + bc$, where the first term in each rate represents normal cell death without birth and the second term represents budding. As discussed earlier, budding can be viewed as a "death" followed immediately by birth and regeneration of the parent particle. All viruses die at rate v. The progeny generating functions for the model are

$$\begin{aligned}
P_1(s) &= u + (1-m)(1-u)s_4 + m(1-u)s_5,\\
P_2(s) &= u + (1-m)(1-u)s_5 + m(1-u)s_6,\\
P_3(s) &= u + (1-u)s_6,\\
P_4(s) &= \frac{1}{1+f_4b} + \frac{f_4b}{1+f_4b}s_1s_4,\\
P_5(s) &= \frac{1}{1+f_5b} + \frac{f_5b}{1+f_5b}s_2s_5,\\
P_6(s) &= \frac{1}{1+b} + \frac{b}{1+b}s_3s_6.
\end{aligned}$$

The available data suggest that $v = 2.77$, $c = 0.46$, and $m = 4 \times 10^{-5}$ per day [31, 34]. Because there are little data on the absolute fitness of viral variants during therapy, it is difficult to set the burst size b and the immune clearance rate u. However, data on the rebound rate of resistant mutants during therapy are available [33], and matching this rebound rate to the maximum eigenvalue of the branching process matrix Ω, we estimate $(1-u)b = 2$. For the sake of simplicity, we set $u = 0$ and $b = 2$. Based on the infectivities of the mutants relative to the wild type in the presence of 625 nM ritonavir, we estimate $f_4b = 0.24$ for the single mutant and $f_5b = 0.32$ for the double mutant [30]. To pick starting conditions, we use the deterministic model of Ribeiro, Bonhoeffer, and Nowak [39] and the estimated fitness of mutant viruses relative to wild type viruses in the absence of ritonavir [30]. Assuming 1×10^8 total infected cells at steady state off treatment [9], we estimate

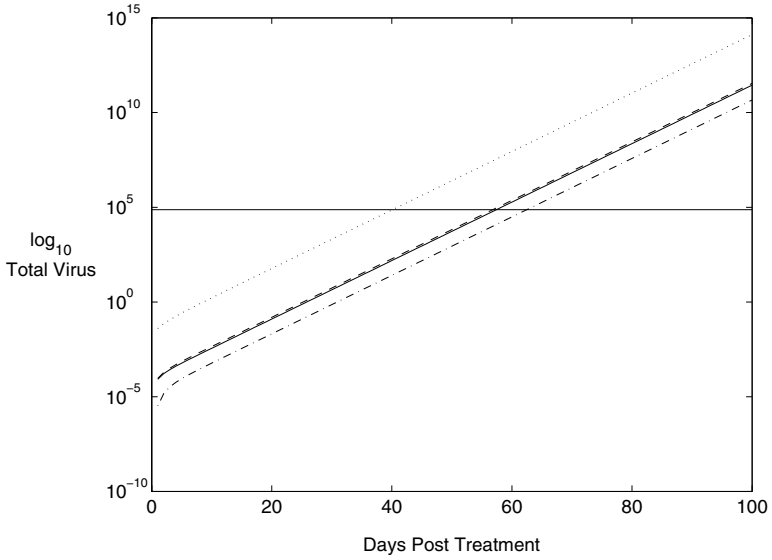

Fig. 5 *HIV model results.*

that there are 2.3×10^4 cells infected with single mutant, 22 cells infected with double mutant, and no cell infected with the triple mutant at steady state prior to treatment. Under the Ribeiro–Bonhoeffer–Nowak model [39], the initial number of type-i viruses is a constant multiple $(1-s_i)bc/v$ of the initial number of corresponding infected cells, where s_i is the selection coefficient against the virus mutant in the absence of drugs. The starting numbers for all particle types are shown in the last column of Table 2.

As a numerical illustration of the model with these parameter values, we find that the mean number of virus particles achieves detectable levels (approximately 25 viruses/ml plasma) after 58 days of therapy. The base 10 logged mean of the number of triple mutant viruses is plotted against time as a solid line in Figure 5. The mean number plus 1.6 times the standard deviation is plotted as a dotted line. The horizontal line is the clinically detectable level. Assuming for the sake of argument that high particle counts are normally distributed, 5% of patients experience detectable levels just 40 days into treatment.

We have ignored other mutational pathways to the triple mutant. In fact, the estimates of pretreatment fitness of the 54V mutant virus are higher than that for the 82A mutant virus, indicating that there should be more 54V mutants present before therapy starts. Despite this observation, the 54V mutation is the last to appear in treated patients [33]. The reason for the discrepancy is likely to be the low on-treatment fitness level of the dual mutants involving 54V. Unfortunately, Mammano et al. [30] did not test the dual mutant 71V54V. If we assume the fitness of 71V54V to be the same as 82A54V and force the virus to obtain resistance starting with the 54V single mutant (dash-dot paths in Figure 4), complete resistance would be achieved an average 5 days later, as shown by the dash-dot line in Figure 5. Considering all possible pathways to resistance does not substantially speed up the appearance of the triple mutant as shown by the dashed line in Figure 5. Whatever the most likely sequence of mutations on the way to full resistance, it is clear that the stochastic nature of mutation explains much of the variation observed between patients on ritonavir treatment.

13. Conclusion. With their emphasis on growth, reproduction, immigration, and extinction, the biological sciences offer fertile terrain for the application of branching processes. Although it is easy to formulate meaningful models, it is difficult to manipulate them mathematically and draw relevant conclusions. Branching process theory is dominated by a handful of tractable models and an elaborate asymptotic machinery. Except for brute force simulation, the middle way of numerical analysis has been little explored. Here we have surveyed and extended numerical methods for Markovian multitype processes in continuous time. The mix of techniques from linear algebra, ordinary differential equations, probability theory, and Fourier analysis is both attractive and effective.

This is not to imply that the techniques are perfect or that we possess a rigorous mathematical basis for predicting their numerical accuracy. What evidence we have on accuracy is reassuring. In the haplotype model, our current techniques produce estimates for the means and variances of type counts that agree well with previous estimates based on Laplace transforms [8, 26] and with direct estimates from the marginal distributions shown in Figure 2. The later comparisons inspire confidence in our computationally intensive method of approximating the marginal distributions. This method, which combines Fourier inversion with numerical integration of the backward equations, can be implemented in commercially available software such as MATLAB.

We cannot stress enough the value of modeling immigration as a Poisson process. Our applications to genetics, disease spread, and virus population dynamics reinforce this point. Poisson processes bring into play Campbell's theorem and the potent extension of Poisson processes by marking and coloring. Of course, the assumptions of independent Poisson streams of immigrants are not universally satisfied. For example, it would be rewarding to add a therapeutic sanctuary as another source of partially resistant virus in the HIV model. Modeling this source as independent Poisson immigration is at best a crude approximation, given our suspicion that viruses evolving in the presence of drugs can migrate back to the sanctuary.

Despite the limitations of branching processes, we believe the time is ripe for a revival. Branching process models give a much richer picture than their deterministic counterparts. Phenomena such as extinction are inherently stochastic. Whenever small numbers and stochastic fluctuations prevail in biological models, deterministic approaches are fatally handicapped. Mathematical scientists are fortunate to have an alternative and should exploit it accordingly.

Appendix. We seek the variance matrix $V_i(t)$ for the process starting with one particle of type i at time 0. Conditioning on the state of the process at time $t < t+\tau$, we infer that

$$\begin{aligned} V_i(t+\tau) &= \mathrm{Var}[\mathrm{E}\,(Z_{t+\tau} \mid Z_t)] + \mathrm{E}\,[\mathrm{Var}(Z_{t+\tau} \mid Z_t)] \\ &= e^{\tau\Omega^*} V_i(t) e^{\tau\Omega} + \mathrm{E}\,[\mathrm{Var}(Z_{t+\tau} \mid Z_t)] \end{aligned} \tag{A.1}$$

since $\mathrm{E}\,(Z_{t+\tau} \mid Z_t) = Z_t e^{\tau\Omega}$ and $(e^{\tau\Omega})^* = e^{\tau\Omega^*}$. To simplify (A.1), we invoke formula (17). This yields

$$\mathrm{Vec}[V_i(t+\tau) - V_i(t)] = \left[e^{\tau\Omega^*} \otimes e^{\tau\Omega^*} - I \otimes I\right] \mathrm{Vec}[V_i(t)] + \mathrm{Vec}[\Upsilon_i(t,\tau)],$$

where $\Upsilon_i(t,\tau) = \mathrm{E}\left[\mathrm{Var}(Z_{t+\tau} \mid Z_t)\right]$ and $I \otimes I$ is the $r^2 \times r^2$ identity matrix. The product rule of elementary calculus gives

$$\begin{aligned}\lim_{\tau\to 0} \frac{1}{\tau}\left(e^{\tau\Omega^*} \otimes e^{\tau\Omega^*} - I \otimes I\right) &= \frac{d}{d\tau}\left(e^{\tau\Omega^*} \otimes e^{\tau\Omega^*}\right)\Big|_{\tau=0} \\ &= \left(\Omega^* e^{\tau\Omega^*} \otimes e^{\tau\Omega^*} + e^{\tau\Omega^*} \otimes e^{\tau\Omega^*}\Omega^*\right)\Big|_{\tau=0} \\ &= \Omega^* \otimes I + I \otimes \Omega^*,\end{aligned}$$

and therefore

$$\frac{d}{dt}\,\mathrm{Vec}[V_i(t)] = (\Omega^* \otimes I + I \otimes \Omega^*)\,\mathrm{Vec}[V_i(t)] + \lim_{\tau\to 0}\frac{1}{\tau}\,\mathrm{Vec}[\Upsilon_i(t,\tau)].$$

To find $\lim_{\tau\to 0}\tau^{-1}\Upsilon_i(t,\tau)$, let $m_{ij}(t)$ denote the expected number of particles of type j starting from a single particle of type i. Because different clans evolve independently,

$$\mathrm{E}\left[\mathrm{Var}(Z_{t+\tau} \mid Z_t)\right] \;=\; \sum_{j=1}^{r} m_{ij}(t)V_j(\tau) \;=\; \sum_{j=1}^{r} m_{ij}(t)\left[V_j(\tau) - V_j(0)\right].$$

This fact implies that

$$\lim_{\tau\to 0}\frac{1}{\tau}\Upsilon_i(t,\tau) = \sum_{j=1}^{r} m_{ij}(t)\frac{d}{dt}V_j(0)$$

and consequently that

$$\frac{d}{dt}\,\mathrm{Vec}[V_i(t)] = (\Omega^* \otimes I + I \otimes \Omega^*)\,\mathrm{Vec}[V_i(t)] + \sum_{j=1}^{r} m_{ij}(t)\,\mathrm{Vec}(C_j) \tag{A.2}$$

for $C_j = \frac{d}{dt}V_j(0)$.

Further simplification can be achieved by stacking the vectors $\mathrm{Vec}[V_i(t)]$ and $\mathrm{Vec}(C_i)$ into vectors $\mathrm{Vec}[V(t)]$ and $\mathrm{Vec}(C)$ and substituting $e^{t\Omega}$ for $[m_{ij}(t)]$. In view of the identities (18) and (19), the stacked version of (A.2) amounts to nothing more than

$$\frac{d}{dt}\,\mathrm{Vec}[V(t)] = I \otimes (\Omega^* \otimes I + I \otimes \Omega^*)\,\mathrm{Vec}[V(t)] + \left(e^{t\Omega} \otimes I \otimes I\right)\mathrm{Vec}(C).$$

Because $V(0)$ vanishes, this constant-coefficient, linear differential equation has solution

$$\mathrm{Vec}[V(t)] = \int_0^t e^{(t-\tau)I\otimes(\Omega^*\otimes I + I\otimes\Omega^*)}e^{\tau\Omega} \otimes I \otimes I d\tau\,\mathrm{Vec}(C).$$

Using the commutativity of $I \otimes \Omega^* \otimes I$ and $I \otimes I \otimes \Omega^*$ and the series definition of the matrix exponential, one can easily show that

$$e^{(t-\tau)I\otimes(\Omega^*\otimes I + I\otimes\Omega^*)} = I \otimes e^{(t-\tau)\Omega^*} \otimes e^{(t-\tau)\Omega^*}.$$

It follows that

$$\mathrm{Vec}[V(t)] = \int_0^t e^{\tau\Omega} \otimes e^{(t-\tau)\Omega^*} \otimes e^{(t-\tau)\Omega^*} d\tau\,\mathrm{Vec}(C). \tag{A.3}$$

Formula (A.3) still lacks a specific value of C. To determine one of the matrices $C_i = \frac{d}{dt}V_i(0)$ comprising it, let

$$\begin{aligned} f_i &= dP_i(\mathbf{1}), \\ G_i &= d^2P_i(\mathbf{1}) + D(f_i) - f_i^* f_i \end{aligned}$$

be the mean vector and variance matrix of the progeny generating function $P_i(s)$. Here $dP_i(s)$ is the first differential (row vector of first partials), $d^2P_i(s)$ is the second differential (Hessian matrix of second partials), and $D(f_i)$ is the diagonal matrix with jth diagonal entry f_{ij}. If X is the indicator of the random event that the founding type-i particle dies during $[0,t]$, then the usual conditioning argument gives

$$V_i(t) = \mathrm{E}\,[\mathrm{Var}(Z_t \mid X)] + \mathrm{Var}[\mathrm{E}\,(Z_t \mid X)].$$

Because at most one event occurs during $[0,t]$ for t small, we have

$$\begin{aligned} \mathrm{E}\,[\mathrm{Var}(Z_t \mid X)] &= (1-\omega_i t)\,\mathrm{Var}(Z_t \mid X=0) + \omega_i t\,\mathrm{Var}(Z_t \mid X=1) + o(t) \\ &= \omega_i t G_i + o(t). \end{aligned}$$

Similarly, if u_i is the standard unit vector, then

$$\begin{aligned} \mathrm{Var}[\mathrm{E}\,(Z_t \mid X)] &= \mathrm{E}\,[\mathrm{E}\,(Z_t \mid X)^*\,\mathrm{E}\,(Z_t \mid X)] - \mathrm{E}\,(Z_t)^*\,\mathrm{E}\,(Z_t) \\ &= (1-\omega_i t)u_i^* u_i + \omega_i t f_i^* f_i \\ &\quad - [(1-\omega_i t)u_i + \omega_i t f_i]^*[(1-\omega_i t)u_i + \omega_i t f_i] + o(t) \\ &= \omega_i t(u_i^* u_i + f_i^* f_i - u_i^* f_i - f_i^* u_i) + o(t). \end{aligned}$$

From these considerations, it follows that

$$C_i \;=\; \frac{d}{dt}V_i(0) \;=\; \omega_i(G_i + u_i^* u_i + f_i^* f_i - u_i^* f_i - f_i^* u_i). \tag{A.4}$$

When the matrix Ω is diagonalizable, formula (A.3) can be explicitly evaluated. If in obvious notation $\Omega = A\Delta A^{-1}$, then

$$\mathrm{Vec}[V(t)] = H\int_0^t e^{\tau\Delta} \otimes e^{(t-\tau)\Delta} \otimes e^{(t-\tau)\Delta} d\tau H^{-1}\,\mathrm{Vec}(C),$$

where $H = A \otimes (A^*)^{-1} \otimes (A^*)^{-1}$ and $H^{-1} = A^{-1} \otimes A^* \otimes A^*$. The matrix integral $\int_0^t e^{\tau\Delta} \otimes e^{(t-\tau)\Delta} \otimes e^{(t-\tau)\Delta} d\tau$ is diagonal. At level (i,j,k) its diagonal entry is

$$\int_0^t e^{\tau\delta_i} e^{(t-\tau)\delta_j} e^{(t-\tau)\delta_k} d\tau = \frac{1}{\delta_i - \delta_j - \delta_k}\left[e^{\delta_i t} - e^{(\delta_j+\delta_k)t}\right], \tag{A.5}$$

where δ_i is the ith diagonal entry of Δ.

REFERENCES

[1] S. Asmussen and H. Hering, *Branching Processes*, Birkhäuser, Boston, 1983.
[2] K. B. Athreya and P. E. Ney, *Branching Processes*, Springer-Verlag, New York, 1972.
[3] M. S. Bartlett, *An Introduction to Stochastic Population Models with Special Reference to Methods and Applications*, Cambridge University Press, Cambridge, UK, 1955, p. 77.
[4] R. Bartoszynski, *On a certain model of an epidemic*, Zastos. Mat., 13 (1972), pp. 139–151.
[5] N. G. Becker, *Analysis of Infectious Disease Data*, Chapman and Hall, New York, 1989.

[6] N. G. Becker and K. Dietz, *Reproduction numbers and critical immunity levels for epidemics in a community of households*, in Athens Conference on Applied Probability and Time Series Analysis, Volume I: Applied Probability, C. C. Heyde, Y. V. Prohorov, R. Pyke, and S. T. Rachev, eds., Springer-Verlag, New York, 1996, pp. 267–276.
[7] N. G. Becker and J. L. Hopper, *Assessing the heterogeneity of disease spread through a community*, Amer. J. Epidemiol., 117 (1983), pp. 362–374.
[8] R. Fan and K. Lange, *Models for haplotype evolution in a nonstationary population*, Theor. Popul. Biol., 53 (1998), pp. 184–198.
[9] T. W. Chun, L. Carruth, D. Finzi, X. Shen, J. A. DiGiusepp, H. Taylor, M. Hermankova, K. Chadwick, J. Margolick, T. C. Quinn, Y. -H. Kuo, R. Brookmeyer, M. A. Zeiger, P. Barditch-Cro, and R. F. Siliciano, *Quantification of latent tissue reservoirs and total body viral load in HIV-*1 *infection*, Nature, 387 (1997), pp. 183–188.
[10] K. S. Crump and C. J. Mode, *A general age-dependent branching process*, J. Math. Anal. Appl., 24 (1968), pp. 494–508.
[11] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd ed., Wiley, New York, 1968, pp. 275–276, 287.
[12] K. Gladstien and K. Lange, *Equilibrium distributions for deleterious genes in large stationary populations*, Theor. Popul. Biol., 14 (1978), pp. 322–328.
[13] K. Gladstien and K. Lange, *Number of people and number of generations affected by a single deleterious mutation*, Theor. Popul. Biol., 14 (1978), pp. 313–321.
[14] T. E. Harris, *The Theory of Branching Processes*, Dover, New York, 1989, p. 99.
[15] J. Hästbacka, A. de la Chapelle, I. Kaitila, P. Sistonen, A. Weaver, and E. Lander, *Linkage disequilibrium mapping in isolated founder populations: Diastrophic dysplasia in Finland*, Nat. Genet., 2 (1992), pp. 204–211.
[16] P. Henrici, *Fast Fourier methods in computational complex analysis*, SIAM Rev., 21 (1979), pp. 481–527.
[17] M. W. Hirsch and S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra*, Academic Press, New York, 1974, pp. 192–198.
[18] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1991.
[19] J. A. Jacquez, C. P. Simon, and J. S. Koopman, *The reproduction number in deterministic models of contagious diseases*, Comments Theor. Biol., 2 (1991), pp. 159–209.
[20] P. Jagers, *Branching Processes with Biological Applications*, Wiley, New York, 1975.
[21] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, 2nd ed., Academic Press, New York, 1975, pp. 412–431.
[22] M. Kimmel and D. E. Axelrod, *Branching Processes in Biology*, Springer-Verlag, New York, 2002.
[23] J. F. C. Kingman, *Poisson Processes*, Clarendon Press, Oxford, 1993.
[24] K. Lange, *Calculation of the equilibrium distribution for a deleterious gene by the finite Fourier transform*, Biometrics, 38 (1982), pp. 79–86.
[25] K. Lange, M. Boehnke, and R. Carson, *Moment computations for subcritical branching processes*, J. Appl. Probab., 18 (1981), pp. 52–64.
[26] K. Lange and R. Fan, *Branching process models for mutant genes in nonstationary populations*, Theor. Popul. Biol., 51 (1997), pp. 118–133.
[27] K. Lange and K. Gladstien, *Further characterization of the long-run population distribution of a deleterious gene*, Theor. Popul. Biol., 18 (1980), pp. 31-43.
[28] C. A. Macken and A. S. Perelson, *Branching Processes Applied to Cell Surface Aggregation Phenomena*, Lecture Notes in Biomath. 58, Springer-Verlag, New York, 1985.
[29] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, New York, 1999, pp. 30–31, 158–165.
[30] F. Mammano, V. Trouplin, V. Zennou, and F. Clavel, *Retracing the evolutionary pathways of human immunodeficiency virus type* 1 *resistance to protease inhibitors: Virus fitness in the absence and in the presence of drug*, J. Virol., 74 (2000), pp. 8524–8531.
[31] L. M. Mansky, *Forward mutation rate of human immunodeficiency virus type* 1 *in a T lymphoid cell line*, AIDS Res. Hum. Retroviruses, 12 (1996), pp. 307–314.
[32] C. J. Mode, *Multitype Branching Processes: Theory and Applications*, American Elsevier, New York, 1971.
[33] A. Molla, M. Korneyeva, Q. Gao, S. Vasavanonda, P. J. Schipper, H. M. Mo, M. Markowitz, T. Chernyavskiy, P. Niu, N. Lyons, A. Hsu, G. R. Granneman, D. D. Ho, C. A. B. Boucher, J. M. Leonard, D. W. Norbeck, and D. J. Kempf, *Ordered accumulation of mutations in HIV protease confers resistance to ritonavir*, Nat. Med., 2 (1996), pp. 760–766.

[34] A. S. Perelson, P. Essunger, Y. Cao, M. Vesanen, A. Hurley, K. Saksela, M. Markowitz, and D. D. Ho, *Decay characteristics of HIV-1-infected compartments during combination therapy*, Nature, 387 (1997), pp. 188–191.

[35] A. S. Perelson and P. W. Nelson, *Mathematical analysis of HIV-1 dynamics in vivo*, SIAM Rev., 41 (1999), pp. 3–44.

[36] A. N. Phillips, N. Youle, M. Johnson, and C. Loveday, *Use of a stochastic model to develop understanding of the impact of different patterns of antiretroviral drug use on resistance development*, AIDS, 15 (2001), pp. 2211–2220.

[37] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran: The Art of Scientific Computing*, 2nd ed., Cambridge University Press, Cambridge, UK, 1992.

[38] I. Rahimov, *Random Sums and Branching Stochastic Processes*, Springer-Verlag, New York, 1995.

[39] R. M. Ribeiro, S. Bonhoeffer, and M. A. Nowak, *The frequency of resistant mutant virus before antiretroviral therapy*, AIDS, 12 (1998), pp. 461–465.

[40] R. M. Ribeiro and S. Bonhoeffer, *Production of resistant HIV mutants during antiretroviral therapy*, Proc. Natl. Acad. Sci. USA, 97 (2000), pp. 7681–7686.

[41] E. Seneta, *Non-negative Matrices: An Introduction to Theory and Applications*, Wiley, New York, 1973, pp. 40–43.

[42] Z. Taïb, *Branching Process and Neutral Evolution*, Lecture Notes in Biomath. 93, Springer-Verlag, New York, 1992.