Detecting Particular Genotypes in Populations Under Nonrandom Mating*

S. KARLIN
Department of Mathematics, Stanford University, Stanford, California 94305

AND

S. TAVARÉ
Department of Statistics, Colorado State University, Fort Collins, Colorado 80523

Received 1 September 1981

ABSTRACT

We investigate the time to formation of particular genotypes in populations with *nonrandom* mating systems. We employ two main techniques. The first is a study of branching processes with "killing"; these are models which behave just like a standard Galton-Watson branching process with the added possibility of being terminated by the occurrence of a special event in the process. In our case, this special event corresponds to formation or detection of a group of individuals carrying a specific genotype. We then use these results and some natural approximation methods to analyze and interpret the gene formation problem in a simple way.

INTRODUCTION

The basic framework is the following. In a population comprising N individuals, each classified as one of three possible genotypes AA, Aa, aa, how long does it take to produce or form the first aa-genotype under different mating and selection regimes? Under the assumption of a classical Wright-Fisher reproduction scheme, Robertson [11] ascertained by matrix numerical methods and simulation that if the population contained initially one heterozygote (the remaining individuals being AA), then the time to formation (and detection, since the aa-genotype is assumed visible on formation) of the first aa individuals takes about $2N^{1/3}$ generations. By comparision, the time to fixation of the a-allele, given that it is a new mutant destined for fixation, is about $4N$ generations [10]. Robertson's original model and a variety of extensions have been analyzed by the authors [5–7] using the method of diffusion approximation to the underlying Markov-chain models.

*Supported in part by NIH Grant NIH 5R01 GM10452-18 and NSF Grant MCS79-24310.

52 Vanderbilt Ave., New York, NY 10017 0025-5564/82/03057+19$02.75

The novelty in the analysis lies in the use of diffusion processes with killing, this killing term deriving from the appearance of the new visible genotypes. For example, the time to detection of *aa*-types corresponds to the killing time of the diffusion process. Using this approach we were able to confirm and extend Robertson's results, and provide a natural way to analyze formation or detection times.

Implicit in these analyses is the appearance of an approximation process that is continuous both in its time scale and its state space. However, under a variety of reproduction schemes typically involving nonrandom mating or strong selection effects, the approximating process is no longer continuous in this way, but is *discrete* in both time and state space. It is these processes we want to discuss in this paper.

The layout of the paper is as follows. In Section I, we describe a simple model of a branching process with killing corresponding to the detection in the population of individuals that are defective in some way. As will be seen, processes of this sort arise in the later sections as approximations to our genetic models, and so we highlight their behavior at the beginning.

Section II discusses the detection problem in the case of selfing schemes where we assess the effects of selection and incomplete penetrance in heterozygotes on the time to formation of the *aa*-genotype. We also discuss the behavior of a sex-linked trait (cf. [4]). In Section III we focus attention on models which lead to multidimensional branching processes. Among these, we analyze the effects of imperfect (partial) visibility of the *aa*-genotype. We also study sib-mating and parent-offspring-mating systems.

Although our primary interest is in the genetic applications, we include a brief appendix on the nature of the approximation methods used in Section II.

I. BRANCHING PROCESSES WITH KILLING

In this section we will develop briefly some results concerning one-dimensional Galton-Watson branching processes with killing. The results and methods derived here are used to analyze the approximating processes described in Section II. Motivated by these applications, we will use the term *detection* interchangeably with the word *killing*, so no confusion will arise.

We consider a population of individuals that reproduces in the following way. Each individual alive at a particular time produces, independently of the others alive at that time, a random number of offspring, each with the distribution of a random variable Z satisfying

$$\Pr\{Z=k\}=p_k, \quad k\geq 0; \qquad p_0>0, \quad p_0+p_1<1. \tag{1}$$

An offspring born to a particular individual has a probability $1-\alpha$ of being

found to be defective in some way. We assume in this simplest case that detection of defectives is independent in a family, and over families. To avoid trivialities, we also will assume that $0<\alpha<1$. A family of size k survives to reproduce if, and only if, all k individuals are normal. This has probability $p_k\alpha^k, k\geq 0$. It follows that if $f(s)=\sum_{k\geq 0}p_k s^k$ is the p.g.f. of Z, then the (improper) p.g.f. of the number of offspring born to an individual with no defective offspring is

$$g(s)=f(\alpha s), \tag{2}$$

and $1-g(1)=1-f(\alpha)$ is the probability that a family contains any defective individuals. The population now evolves as follows. Let X_n be the number of individuals alive at time n. The population continues to the next generation only if no defective individuals are born. Otherwise, we say the process has ended by a killing (or detection) event. Under the simple detection scheme introduced above, it is in principle straightforward to analyze the process. We take $X_0=1$, and define the iterates of $g(\cdot)$ by

$$g_0(s)=s, \qquad g_n(s)=g(g_{n-1}(s)), \quad n\geq 1. \tag{3}$$

Intuitively, it is clear that the process ends either in extinction or in detection. Let q be the probability that extinction prevails. It is simple to show that if $0<\alpha<1$, then q is the unique root satisfying $0<q<1$ of the functional equation $g(s)=f(\alpha s)=s$; if $X_0=i>1$, then the extinction probability is q^i.

The probability that the detection time T_D is greater than n is given by

$$\Pr\{T_D>n\}=\Pr\{0\leq X_n\}=g_n(1).$$

Since $g_n(s)$ is decreasing in n for $s\in(q,s_1)$, where s_1 is the larger root of $f(\alpha s)=s$ satisfying $s_1>1$, we conclude that $g_n(1)\to q$ as $n\to\infty$. This establishes that indeed the process terminates either by detection or by extinction. The probability that detection prevails is then $1-q^i$ if $X_0=i$.

Two relevant distributions in the study of detection times are the (conditional) detection time T_D, and $T=\min(T_0,T_D)$, the time to extinction or detection. We have

$$\Pr[T_D>n\,|\,T_D<\infty]=\frac{g_n(1)-q}{1-q}, \qquad n\geq 0, \tag{4}$$

and

$$\Pr[T>n]=g_n(1)-g_n(0), \qquad n\geq 0. \tag{5}$$

We determine the asymptotic behavior of (4) and (5) as follows. Let $\gamma=g'(q)=\alpha f'(\alpha q)<1$. A modification of the proof provided by Athreya

and Ney [1, pp. 38–41] establishes the existence of

$$Q(s)=\lim_{n\to\infty}\gamma^{-n}(g_n(s)-q),\qquad 0\leqslant s<s_1, \tag{6}$$

where $Q(s)$ satisfies the functional equation $Q(g(s))=\gamma Q(s)$, subject to $Q(q)=0$, $Q'(q)=1$, and $Q'(s)>0$ for every $s\in[0,s_1)$. It follows immediately that $\Pr[T>n]=O(\gamma^n)$, and $\Pr[T_D>n|T_D<\infty]=O(\gamma^n)$, $n\to\infty$. Hence both T and T_D (conditioned on $T_D<\infty$) have finite moments of all orders. We can use the result in (6) to establish other interesting asymptotic properties of the transition probabilities. We give one example involving the asymptotic conditional distribution.

Fix $X_0=1$, and set $a_j^{(n)}=\Pr[X_n=j|T>n]$ and define $\varphi^{(n)}(s)=\sum_{j=1}^{\infty}a_j^{(n)}s^j$, $|s|\leqslant 1$. Using (4) and (5), we find that $\varphi^{(n)}(s)=[g_n(s)-g_n(0)]/[g_n(1)-g_n(0)]$. Now use (6) to see that

$$\varphi^{(n)}(s)=\frac{\gamma^{-n}(g_n(s)-q)+\gamma^{-n}(q-g_n(0))}{\gamma^{-n}(g_n(1)-q)+\gamma^{-n}(q-g_n(0))}\to\frac{Q(s)-Q(0)}{Q(1)-Q(0)}\qquad\text{as}\quad n\to\infty. \tag{7}$$

The right-hand side of (7) is the probability generating function of the asymptotic conditonal distribution $a_j=\lim_{n\to\infty}a_j^{(n)}$.

Loosely speaking we interpret a_j as follows. If the process has been running for a long time, and neither detection nor extinction has occurred, then X is in state j with probability a_j. The mean of this asymptotic conditional distribution is $\sum_{j\geqslant 1}ja_j=Q'(1)/[Q(1)-Q(0)]$.

There are essentially only two cases where explicit forms for the iterates of $g(\cdot)$ are available. One is the trivial case $f(s)=p+(1-p)s$, $0<p<1$, and the other is the linear fractional p.g.f.

$$f(s)=\frac{r+s(1-r-p)}{1-ps},\qquad 0<r<1,\quad 0<p<1,\quad p+r\leqslant 1, \tag{8}$$

which corresponds to

$$p_k=\begin{cases} r, & k=0,\\ (1-r)(1-p)p^{k-1}, & k\geqslant 1.\end{cases} \tag{9}$$

In this latter case, we have

$$g(s)=\frac{r+s\alpha(1-r-p)}{1-p\alpha s}. \tag{10}$$

Let $0<s_0<1<s_1$ be the roots of the equation $g(s)=s$, and define

$$K=\frac{1-p\alpha s_1}{1-p\alpha s_0}. \tag{11}$$

It is readily checked that $0<K<1$, and using a standard method for iterating a linear fractional function (cf. [8, p. 403]), one obtains

$$g_n(s)=\frac{s_0s_1(K^n-1)+s(s_0-K^ns_1)}{(K^ns_0-s_1)-s(K^n-1)}, \qquad 0\leqslant s<s_1. \tag{12}$$

It follows immediately from (11) that $P(T_0\leqslant n)=g_n(0)\to s_0$ as $n\to\infty$, so that $q=P(T_0<\infty)=s_0$. Further, from (4),

$$P[T_D>n|T_D<\infty]=\frac{K^n(s_1-s_0)}{K^n(1-s_0)+(s_1-1)}. \tag{13}$$

In this case, $K=\gamma$, and so the function $Q(\cdot)$ is given, referring to (11) by

$$Q(s)=\frac{(s_1-s_0)(s_0-s)}{(s-s_1)}, \qquad 0\leqslant s<s_1. \tag{14}$$

The asymptotic conditional distribution $\{a_j\}$ specified at (7) has generating function

$$\frac{Q(s)-Q(0)}{Q(1)-Q(0)}=\frac{s(1-s_1^{-1})}{1-s_1^{-1}s}, \qquad 0\leqslant s<1,$$

showing that $\{a_j\}$ is the geometric distribution

$$a_j=(1-s_1^{-1})s_1^{-j}, \qquad j\geqslant 1, \tag{15}$$

with mean $(1-s_1^{-1})^{-1}$.

Another pertinent conditioning focuses attention on those paths that become extinct, rather than those that end in detection. The process $\{\tilde{X}_n, n\geqslant 0\}$ that arises by conditioning on extinction is again a branching process, with offspring p.g.f. given by

$$\tilde{f}(s)=\frac{g(sq)}{q}=\frac{p\alpha s_1+s[1-p\alpha(s_1+s_0)]}{1-sp\alpha s_0}, \qquad 0\leqslant s<1.$$

The process is subcritical (as it must be) with offspring mean $\tilde{f}'(1)=K<1$.

For comparison with the paths allowing both extinction and detection, we will evaluate the asymptotic conditional distribution of $\tilde{X}_n$. Let $\tilde{T}_0$ be the time to hit $\{0\}$. Then it is straightforward to show that if $\tilde{X}_0 = 1$,

$$\tilde{a}_j = \lim_{n \to \infty} P[\tilde{X}_n = j | \tilde{T}_0 > n] = \left(1 - \frac{s_0}{s_1}\right)\left(\frac{s_0}{s_1}\right)^{j-1}, \qquad j \geqslant 1. \tag{16}$$

This distribution has mean $(1 - s_0/s_1)^{-1}$. Comparing this with the result of (15), we see that the mean number of individuals, conditional on the process being "alive", is rather lower in this last case than when detection is also allowed.

2. DETECTING SPECIAL GENES IN FINITE POPULATIONS

2.1. *SELFING SCHEMES*

We consider a diploid population of N individuals, and focus attention on a single locus at which there are two possible alleles, denoted by A and a. To model the evolution of this population, we let X_n denote the number of Aa genotypes in the population at time $n, n = 0, 1, 2, \ldots$. We assume that the population reproduces by selfing, that the population currently comprises only AA and Aa genotypes, and that homozygotes aa are visible (and therefore detected) as soon as they are formed. We want to ascertain the properties of the time taken to detect or form the first aa-individuals. To introduce the methods, and relate them to the results of Section 1, we look at the simplest case first.

Suppose that no aa individuals have been formed up to time n, and that $X_n = i$. Under the selfing scheme the genotypes AA, Aa, aa are produced in proportions q_i, p_i, r_i, where

$$p_i = \frac{i}{2N}, \qquad q_i = 1 - \frac{3i}{4N}, \qquad r_i = \frac{i}{4N}. \tag{17}$$

To form the next generation, we take a multinomial sample of size N according to the probabilities specified in (17). The process continues to the next reproduction stage only if we select no aa-homozygotes, and hence $X_{n+1} = j$ with probability p_{ij}, where

$$p_{ij} = \binom{N}{j} p_i^j q_i^{N-j}, \qquad 0 \leqslant i, j \leqslant N. \tag{18}$$

The transition probabilities in (18) no longer have row sums unity, corresponding to our not considering those outcomes in which a positive number of aa-genotypes are produced. To remedy this, we may add on an extra state

H, say, and define the transition probabilities to H by

$$p_{iH} = 1 - \sum_{j=0}^{N} p_{ij} = 1 - (p_i + q_i)^N,$$

$$p_{HH} = 1, \qquad p_{Hi} = 0, \quad 0 \leqslant i \leqslant N. \tag{19}$$

The process here ends either by hitting the absorbing state at 0 (corresponding to the population comprising only AA individuals), or by hitting the absorbing state H (corresponding to detection of the aa-genotype). Standard Markov chain methods show that the probability u_i that 0 is reached before H starting from $X_0 = i$ satisfies the system of equations

$$u_i = p_{i0} + \sum_{j=1}^{N} p_{ij} u_j, \qquad u_0 = 1. \tag{20}$$

Even for the innocent-looking probabilities specified by (17) and (18) this system is hard to solve explicitly. For small values of N, one could use numerical methods, but for large N this is impractical; we will resort to finding a suitable approximating process that should be good for large values of N. Details of the approximation method may be found in the appendix. Define

$$\varphi_i^{(N)}(s) = \sum_{j=0}^{N} p_{ij} s^j = (p_i s + q_i)^N, \qquad |s| < 1. \tag{21}$$

For the probabilities given in (17) and (18), this reduces to

$$\varphi_i^{(N)}(s) = \left[1 + \frac{i}{N}\left(\frac{s}{2} - \frac{3}{4}\right)\right]^N.$$

Letting $N \to \infty$ gives

$$\varphi_i^{(N)}(s) \to \varphi_i(s) = \exp\left\{i\left(\frac{s}{2} - \frac{3}{4}\right)\right\}. \tag{22}$$

This shows that we may approximate the behavior of $\{X_n, n \geqslant 0\}$ by another Markov chain with transition probabilities given by

$$p_{ij} = \exp\left(-\frac{3i}{4}\right)\left(\frac{i}{2}\right)^j \frac{1}{j!}, \qquad i, j \geqslant 0, \tag{23}$$

and killing probabilities

$$p_{iH} = 1 - \exp\left(-\frac{i}{4}\right), \qquad i \geqslant 0. \tag{24}$$

The form of (22) shows that this process is a Galton-Watson process involving killing as described in Section 1, with a Poisson offspring distribution of the form

$$f(s)=\exp\{-\lambda(1-s)\}, \tag{25}$$

with $\lambda=\frac{3}{4}$ and $\alpha=\frac{2}{3}$. Although explicit formulae are not available for the iterates of $g(s)=\exp\{s/2-\frac{3}{4}\}$, the fixation and detection probabilities are readily found. The root q of $g(s)=s$ lying between 0 and 1 is readily found to be $q\approx 0.6556$, and we conclude that if the process starts with just one heterozygote, then the detection probability is $v=1-q=0.3444$. Starting with i heterozygotes, the detection probability is $1-(0.3444)^i$. These values should be a good approximation to the results for the underlying model even for moderate values of N. Of course, we can use the theory outlined in Section 1 to derive other properties of the detection problem. For example, the mean detection time (conditional on detection occurring) starting with a single heterozygote is given from (4) as

$$M_D(1)=1+\sum_{n=1}^{\infty}\frac{g_n(1)-q}{1-q}, \tag{26}$$

and this can be evaluated numerically to give $M_D(1)\approx 1.54$ generations. We see that *aa*-homozygotes are formed rather rapidly in this case, and that unlike the diffusion case discussed in the introduction, the mean time is effectively independent of the population size N. This observation is confirmed by comparison with matrix numerical results obtained for the (finite) Markov chain specified at (18).

2.2. *THE EFFECT OF SELECTION*

In this section, we discuss the effects of selective differentials among the genotypes on the detection probability. Suppose that the relative fitnesses of genotypes AA and Aa are in the ratio $1:1+\theta$, where $\theta>-1$. We obtain the genotype formation probabilities (17) as

$$\begin{aligned} p_i &= \frac{i(1+\theta)}{2N}\left(1+\frac{i\theta}{N}\right)^{-1}, \\ q_i &= \left(1-\frac{3i}{4N}+\frac{i\theta}{4N}\right)\left(1+\frac{i\theta}{N}\right)^{-1}, \\ r_i &= 1-p_i-q_i, \end{aligned} \tag{27}$$

and the transition probabilities are then given by (18) and (27). Paraphrasing the argument of (21) and (22) shows that the approximating process is a

TABLE 1

Detection Probability $v = 1 - q$
in Large Populations Starting with a Single Heterozygote.

θ[a]	q	v
−0.4	.8140	.1860
−0.1	.6966	.3034
0	.6556	.3444
0.01	.6515	.3585
0.1	.6144	.3856
0.4	.4948	.5052

[a]Selection parameter.

branching process with

$$\varphi_i(s) = \exp\left\{ i(1+\theta)\left(\frac{s}{2} - \frac{3}{4}\right)\right\},$$

so that the underlying branching process has Poisson offspring p.g.f. given by (25) with $\lambda = \frac{3}{4}(1+\theta)$, $\alpha = \frac{2}{3}$. The detection probability is $v = 1 - q$, where q is given by the root of $g(s) = \varphi_1(s) = s$. Numerical results are provided in Table 1 to give an idea of the effects of selection.

2.3. *INCOMPLETE PENETRANCE IN HETEROZYGOTES*

It is conceivable that there is incomplete penetrance in heterozygotes, resulting in some *Aa* heterozyotes being phenotypically identical to the *aa*-genotype. We will assume that any heterozygote has probability β of being phenotypically an *aa*-type, this operating independently for all heterozygotes. We call β the penetrance probability. Our process now terminates *either* by detection of any *aa*-individuals *or* in the detection of an *Aa* individual who is phenotypically *aa*. For simplicity, we will suppose that there are no selective differentials operating in the system. The genotype formation probabilities are now given by

$$p_i = \frac{i}{2N}(1-\beta), \qquad q_i = 1 - \frac{3i}{4N}, \qquad r_i = 1 - p_i - q_i, \tag{28}$$

and the transition probabilities are specified by (28), (18), and (19). A now familiar analysis leads to the branching process approximation with Poisson offspring distribution (25), and with $\alpha = 2(1-\beta)/3$ and $\lambda = \frac{3}{4}$. As should be the case, the continuation probability α here is smaller than in the case $\beta = 0$ (Section 2.1) since we have the additional possibility of ending by detection of a heterozygote. To find the fixation probability q, we solve for the smallest

TABLE 2

Detection Probability v
In large populations
Starting from a Single Heterozygote

β[a]	q	v
0	.6556	.3444
0.01	.6524	.3476
0.1	.6261	.3739
0.5	.5407	.4593
0.9	.4839	.5161
0.99	.4735	.5265

[a] Penetrance probability.

positive root of $s = \exp\{-\frac{3}{4} + (1-\alpha)s/2\}$. Some numerical examples are provided in Table 2.

Table 2 shows that the detection probability is only markedly changed when there is a very high penetrance probability β. Indeed, the probability of detection occurring before fixation of the A-gene exceeds 0.5 only when $\beta > 0.7725$.

2.4. *SEX-LINKED TRAITS*

We would like to assess the effect of sex linkage on the detection probability of a (perhaps lethal) gene in males. We formulate the problem in the following way. Consider a population of N males and N females. Females carry the genotypes AA, Aa, aa, while males are either AY or aY. We suppose that the haplotype aY is the one we are trying to detect. Let X_n be the number of Aa genotypes in the population at time n, and suppose that all males at time n are AY. The process continues to the next time point if the system produces no aY types in males. For reasons that will become apparent later, we will assume there are no aa-genotypes in females. By analyzing the results of possible matings, the transition probability matrix is found to be

$$p_{ij} = \binom{N}{j}\left(\frac{i}{2N}\right)^j\left(1-\frac{i}{2N}\right)^{N-j}\left(1-\frac{i}{2N}\right)^N, \qquad 0 \leqslant i, j \leqslant N, \quad (29)$$

while the detection probability is

$$p_{iH} = 1-\left(1-\frac{i}{2N}\right)^N. \quad (30)$$

This time, H refers to the appearance of the lethal in males, while X_n keeps

track of heterozygotes in females. Of course, if X ever reaches 0, the population of females comprises only A genes, so the aY can never appear. $\{0\}$ is again an absorbing state. We have assumed that there are no aa-genotypes among the females; apart from this keeping the process one-dimensional, original aa-females would disappear in the next generation anyway. Yet again the approximating process is of the type determined by (25), with $\lambda = 1, \alpha = \frac{1}{2}$, and so q, the fixation probability starting from a single female carrier Aa, is the solution less than unity of the equation $s = \exp\{-1 + s/2\}$. The solution is $q = 0.4639$, and so the detection probability is $v = 0.5361$. In this case, there is quite a high chance of the gene appearing in males before it is lost in the females.

If we want to use a screening program which will find with some probability γ (independent for each) carriers in females, then assuming we stop on detection of *either* a female carrier *or* any aY males, the matrix in (29) and (30) becomes

$$p_{ij} = \binom{N}{j}\left[\frac{i}{2N}(1-\gamma)\right]^j\left(1-\frac{i}{2N}\right)^{2N-j}, \qquad 0 \leqslant i, j \leqslant N, \tag{31}$$

$$p_{iH} = 1-\left[\left(1-\frac{i}{2N}\right)\left(1-\frac{\gamma i}{2N}\right)\right]^N, \qquad 0 \leqslant i \leqslant N. \tag{32}$$

The approximating branching process has transition matrix

$$p_{ij} = e^{-i}\left[\frac{i(1-\gamma)}{2}\right]^j\frac{1}{j!}, \qquad i, j \geqslant 0, \tag{33}$$

$p_{iH} = 1 - e^{-(i/2)(1+\gamma)}, i \geqslant 0$, again a special case of (25), with $\lambda = 1, \alpha = (1-\gamma)/2$. In Table 3 we give the detection probabilities v for different values of γ.

TABLE 3

Probability v of Detecting
a Female Carrier Aa or a Male aY
in Large Populations
Starting with a Single Female Heterozygote

γ[a]	q	v
0	.4639	.5361
0.1	.4506	.5494
0.3	.4272	.5728
0.7	.3900	.6100
0.9	.3748	.6252

[a] Probability of finding a carrier female.

For a rather different approach to some aspects of detecting genes in a sex-linked trait, see also James [4].

3. MULTITYPE MODELS

In this section we will analyze several schemes which result in multitype branching-process approximations. Although we have not explicitly set forth the background results as we did for the one-dimensional case in Section 1, their derivation is straightforward.

3.1. INCOMPLETE DETECTION IN HOMOZYGOTES

This example is an extension of the selfing case treated in 2.1. In field experiments using artificial selection schemes, it is likely that the *aa*-genotype is not visible as soon as it is formed, and so the detection process need not end as soon as any *aa*-homozygotes are produced. To model the effects of this, we might proceed as follows. We keep the population at fixed size N in each generation, and let (X_n, Y_n) denote the numbers of *Aa*- and *aa*-genotypes in the population at time n. We suppose that there is a probability β, $0 < \beta \leq 1$, that an *aa*-individual is detected, this acting independently for all such individuals. We say that a killing or detection event occurs if at least one *aa* is found.

Suppose that the population has continued to generation n without detection occurring, and that $X_n = i, Y_n = j (0 \leq i + j \leq N)$. By considering the output of (selfing) matings in the populations, we find that the genotypes *AA*, *Aa*, *aa* are produced in proportions $q : p : r$ respectively, where

$$p = \frac{i}{2N}, \qquad r = \frac{j}{N} + \frac{i}{4N}, \qquad q = 1 - p - r. \tag{34}$$

To produce the next generation of l *Aa*, m *aa* (and thus $N - l - m$ *AA*) we take a multinomial sample of size N according to the probabilities p, q, r at (34). In order to continue without ending in detection, we find that this has probability

$$p_{(i,j)(l,m)} = \binom{N}{l, m} p^l [r(1-\beta)]^m q^{N-l-m}, \tag{35}$$

while the probability of a detection event is

$$p_{(i,j),H} = 1 - (1 - r\beta)^N. \tag{36}$$

The relevant approximating process is now a two-dimensional one, whose transition functions are evaluated as follows. Let

$$\varphi_{(i,j)}^{(N)}(z,w) = \sum_{(l,m)} P_{(i,j),(l,m)} z^l w^m = [q + pz + r(1-\beta)w]^N. \quad (37)$$

Then from (34), we see that as $N \to \infty$, $\varphi_{(i,j)}^{(N)}(z,w) \to \varphi_{(i,j)}(z,w)$, where

$$\varphi_{(i,j)}(z,w) = \exp\left\{\left(j + \frac{i}{4}\right)(1-\beta)w + \frac{iz}{2} - \frac{3i}{4} - j\right\}. \quad (38)$$

The detection probability is then $p_{(i,j),H} = 1 - \exp\{-\beta(j + i/4)\}$. Equation (38) identifies the process as a (Poisson-generated) branching process with killing. To find the fixation probability at (0,0), which corresponds to a population comprising only AA genotypes, we use a modification of the method of Section 2. Let (z_0, w_0) be the unique solution of the pair of functional equations

$$\varphi_{(1,0)}(z,w) = z, \qquad \varphi_{(0,1)}(z,w) = w. \quad (39)$$

satisfying in our case $0 < z_0, w_0 < 1$. Then the probability that fixation obtains before detection, starting from an initial configuration (i, j), is $z_0^i w_0^j$, and so the detection probability is $q_{(i,j)} = 1 - z_0^i w_0^j$, $i, j \geq 0$.

In the present case, the equations (39) are

$$\exp\{\tfrac{1}{4}(1-\beta)w + \tfrac{1}{2}z - \tfrac{3}{4}\} = z$$

and

$$\exp\{(1-\beta)w - 1\} = w.$$

In Table 4, we give the value of (z_0, w_0) for different values of β.

For comparison with the one-dimensional case analyzed in Section 2.1, notice that the detection probability starting from i heterozygotes and no

TABLE 4

Fixation Probabilities (z_0, w_0)

β[a]	z_0	w_0
.1	.8349	.6759
.5	.7161	.4639
.7	.6872	.4169
.9	.6650	.3822

[a] Homozygote visibility probability.

aa-individuals is $q_{(i,0)} = 1 - z_0^i$. For the case $\beta = 1$ of Section 1.1, we obtained $z_0 = 0.6556$. We remark that the same methods may be used to analyze the effects of selection, mutation, and incomplete penetrance in heterozygotes on this system.

3.2. SIB-MATING SYSTEMS

A variety of other interesting processes arise in the study of gene formation times in inbreeding systems. In this first case, we will consider a sib-mating system. Consider a population of a constant size of N individuals. Instead of recording the number of particular genotypes in the group, we use a device due to Bartlett and Haldane [2] (see also Karlin [9, p. 291]). We now keep track of the number of mating types among the sibs. The process will terminate either by production of a homozygous *aa*-individual, or by fixation of the mating type $AA \times AA$, which results in fixation of the A-gene. Let (X_n, Y_n) be the numbers of mating types $AA \times Aa, Aa \times Aa$ in the set of N sibs at generation n. We do not need to record other mating types, because the population is assumed to contain no *aa*-individuals. The transition matrix is most expeditiously computed from the following mating table:

		Proportion of offspring mating types		
		$AA \times AA$	$AA \times Aa$	$Aa \times Aa$
Parent	$AA \times AA$	1	0	0
mating	$AA \times Aa$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
types	$Aa \times Aa$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{4}$

Given that $X_n = i, Y_n = j$, mating types $AA \times AA, AA \times Aa, Aa \times Aa$ are produced with probabilities $q : p : r$, respectively, where

$$p = \frac{i}{2N} + \frac{j}{4N}, \qquad q = 1 - \frac{3i}{4N} - \frac{15j}{16N}, \qquad r = \frac{i}{4N} + \frac{j}{4N}, \tag{40}$$

as follows from the mating table. For example,

$$p = \left(1 - \frac{i}{N} - \frac{j}{N}\right) \cdot 1 + \frac{i}{N} \cdot \frac{1}{4} + \frac{j}{N} \cdot \frac{1}{16} = \frac{i}{2N} + \frac{j}{4N}.$$

The subsequent configuration of mating types is found by taking a multinomial sample of size N according to the probabilities p, q, r at (40). Hence,

$$p_{(i,j),(l,m)} = \binom{N}{l, m} q^{N-l-m} r^m p^l, \tag{41}$$

and

$$p_{(i,j),H} = 1 - \left(1 - \frac{7j}{16N}\right)^N. \tag{42}$$

The probability in (42) depends only on j, since aa-homozygotes can only be formed from the $Aa \times Aa$ matings. The recipe of the previous section shows that the generating functions $\varphi_{(i,j)}(z,w)$ is given by

$$\varphi_{(i,j)}(z,w) = \exp\left\{-i\left(\frac{3}{4} - \frac{z}{2} - \frac{w}{4}\right)\right\} \exp\left\{-j\left(\frac{15}{16} - \frac{z}{4} - \frac{w}{4}\right)\right\}, \quad (43)$$

and then the corresponding functional equations (39) have the solution $(z_0, w_0) = (0.8147, 0.5510)$. If the population starts with just a single a-allele, perhaps corresponding to a new mutation, then the detection probability is $1 - q_{(1,0)} = 0.1853$.

3.3. PARENT-OFFSPRING MATING

The final case we examine is a repeated parent-offspring mating system. Again, we keep track of mating types in the group of N matings. Let (X_n, Y_n, Z_n) denote the number of $AA \times Aa$, $Aa \times AA$, $Aa \times Aa$ matings, the first element being the parent genotype, the second the offspring genotypes. Since no aa-individuals have yet appeared, the number of $AA \times AA$ matings is $N - X_n - Y_n - Z_n$. The mating table is now given by

		Parent-offspring mating class in next generation			
		$AA \times AA$	$AA \times Aa$	$Aa \times AA$	$Aa \times Aa$
Parent-offspring	$AA \times AA$	1	0	0	0
mating class	$AA \times Aa$	0	0	$\frac{1}{2}$	$\frac{1}{2}$
in current	$Aa \times AA$	$\frac{1}{2}$	$\frac{1}{2}$	0	0
generation	$Aa \times Aa$	0	0	$\frac{1}{4}$	$\frac{1}{2}$

Given $(X_n, Y_n, Z_n) = (i, j, k)$, the mating types appear in proportions

$$\begin{aligned} AA \times AA &: p_1 = 1 - \frac{i}{N} - \frac{j}{2N} - \frac{k}{N}, \\ AA \times Aa &: p_2 = \frac{j}{2N}, \\ Aa \times AA &: p_3 = \frac{i}{2N} + \frac{k}{4N}, \\ Aa \times Aa &: p_4 = \frac{i}{2N} + \frac{k}{2N}, \end{aligned} \quad (44)$$

and the transition probabilities are

$$p_{(i,j,k),(i',j',k')} = \binom{N}{i', j', k'} p_1^{N-i'-j'-k'} p_2^{i'} p_3^{j'} p_4^{k'},$$

with

$$p_{(i,j,k),H} = 1 - \left(1 - \frac{k}{4N}\right)^N.$$

The generating function is given by

$$\varphi^{(N)}_{(i,j,k)}(z,w,u) = \sum_{(i',j',k')} p_{(i,j,k),(i',j',k')} z^{i'} w^{j'} u^{k'},$$

and it is straightforward to see that the limiting generating function is

$$\varphi_{(i,j,k)}(z,w,u) = \exp\left\{ -i\left(1-\frac{w}{2}-\frac{u}{2}\right) - j\left(\frac{1}{2}-\frac{z}{2}\right) - k\left(1-\frac{w}{4}-\frac{u}{2}\right)\right\}.$$

A routine development establishes that the probability of fixation of the A-gene before detection, i.e., reaching $(0,0,0)$ before H from an initial configuration (i,j,k) is given by $z_0^i w_0^j u_0^k$, where (z_0, w_0, u_0) is the appropriate solution of the functional equations

$$\varphi_{(1,0,0)}(z,w,u) = z, \qquad \varphi_{(0,1,0)}(z,w,u) = w, \qquad \varphi_{(0,0,1)}(z,w,u) = u.$$

Numerical evaluation yields the solution (0.7909, 0.9012, 0.6321). Starting from a single heterozygote present in a parent, i.e., from state $(0,1,0)$, the detection probability is given by $1-0.9012 = 0.0988$.

4. DISCUSSION

In this paper, we have analyzed the properties of a class of branching processes with killing that arise in the study of the process of gene formation in nonrandom mating systems. The study is relevant in artificial selection practices, as it pertains to early detection of carriers of deleterious genes. In evolutionary theory, this problem concerns the elapsed time to observation of a new mutant homozygote type. This problem also bears on objectives of medical genetic screening that attempts to identify carriers of defective genes or chromosomal anomalies. The mating patterns studied include cases of selfing lines and sib-mating and parent-offspring inbreeding practices. This framework also is appropriate for models with sex-linked traits. Several of the branching processes are multidimensional, and we used these to assess the effects of imperfect identification of recessive homozygotes.

Many of the approximations involved Poisson "offspring" distributions. This is a result of our use of the clasical Wright-Fisher, or multinomial, sampling schemes. Other formulations of the reproduction mechanism will lead to non-Poisson reproduction behavior, but of course the same methods apply in their analysis.

The framework of Markov models with killing seems to be a natural one within which to analyze the important question of gene formation times. For some examples in which the approximating process is a diffusion with killing, the problem is also tractable; see [5–7]. We concentrated primary attention on detection probabilities. Moments of the detection or fixation times are accessible by the same theory, and we cited some examples in the text.

APPENDIX. CONVERGENCE TO BRANCHING-KILLING PROCESS

We will prove the convergence of the family of Markov chains described in (18)–(19), as $N \to \infty$, to the branching process (with killing) whose progeny p.g.f. is given in (22). To this end, we need to prove that for each i, and integer r

$$\lim_{N\to\infty} \sum_{j=0}^{N} \tilde{P}_{ij}^{(r)}(N)s^j = \sum_{j=0}^{\infty} P_{ij}^{*(r)}s^j, \qquad 0 \leqslant s \leqslant 1, \tag{45}$$

where $\tilde{P}_{ij}^{(r)}(N)$ is the r-step transition probability matrix induced by (18) with $p_i = i/2N$, $q_i = 1 - 3i/4N$, and $P_{ij}^{*(r)}$ is the r-step transition probability of the branching process with individual (improper) p.g.f.,

$$\varphi(s) = e^{s/2-3/4} \quad \text{and} \quad \sum_{j=0}^{\infty} P_{ij}^{*(r)}s^j = \left[\varphi_r(s)\right]^i = \left[\varphi_{r-1}(\varphi(s))\right]^i. \tag{46}$$

For $r=1$ we immediately have $\Sigma_{j=0}^{N}\tilde{P}_{ij}(N)s^j = (1-(\frac{3}{4}-s/2)i/N)^N$, which clearly converges uniformly over $0 \leqslant s \leqslant 1$ and for each fixed i to $[\varphi(s)]^i$. It follows, as all terms are nonnegative, that for $0 \leqslant s \leqslant 1$,

$$\sum_{j=0}^{\infty} \left|\tilde{P}_{ij}(N) - P_{ij}^{*}\right| s^j \to 0. \tag{47}$$

LEMMA 1

For all $0 \leqslant s \leqslant 1$ *and all integer* r

$$\sum_{j=0}^{N} \tilde{P}_{ij}^{(r)}(N)s^j \leqslant \left[\varphi_r(s)\right]^i. \tag{48}$$

Proof. (by induction on r). For $r=1$ the result follows immediately relying on the inequality $(1-a/N)^N \leqslant e^{-a}$, valid for all a, $-1 \leqslant a \leqslant 1$. Consider now $r \geqslant 2$; then $\tilde{P}_{ij}^{(r)} = \Sigma_{k=0}^{N}\tilde{P}_{ik}\tilde{P}_{kj}^{(r-1)}$ and the induction hypothesis joined with the result for $r=1$ gives $\Sigma_{j=0}^{N}\tilde{P}_{ij}^{(r)}s^j = \Sigma_{k=0}^{N}\tilde{P}_{ik}\Sigma_{j=0}^{N}\tilde{P}_{kj}^{(r-1)}s^j \leqslant \Sigma_{k=0}^{N}\tilde{P}_{ik}[\varphi_{r-1}(s)]^k \leqslant [\varphi(\varphi_{r-1}(s)]^i = [\varphi_r(s)]^i$.

We next prove

LEMMA 2

For fixed i *and uniformly in* $0 \leqslant s \leqslant 1$ *we have*

$$\varliminf_{N\to\infty} \sum_{j=0}^{N} \tilde{P}_{ij}^{(2)}(N)s^j \geqslant \left[\varphi_2(s)\right]^i. \tag{49}$$

Proof. With i fixed we can choose an integer K such that $\sum_{j=K+1}^{\infty} P_{ij}^{*} s^j \leqslant \varepsilon$, and then from (47) we deduce for N large enough that $\sum_{j=K+1}^{N} \tilde{P}_{ij}(N) s^j \leqslant 2\varepsilon$. Observe next that $\sum_{j=0}^{N} P_{ij}^{(2)}(N) s^j \geqslant \sum_{k=0}^{K} \tilde{P}_{ik} \sum_{j=0}^{N} \tilde{P}_{kj} s^j = \sum_{k=0}^{K} \tilde{P}_{ik}[1-(\frac{3}{4}-s/2)k/N]^N$. Now for $0 \leqslant k \leqslant K$ we can determine an integer l depending only on K such that $[1-(\frac{3}{4}-s/2)k/N]^{N-l} \geqslant [\varphi(s)]^k$ for all $0 \leqslant k \leqslant K$. Thus, $\sum_{k=0}^{K} \tilde{P}_{ik}[1-(\frac{3}{4}-s/2)k/N]^N \geqslant \sum_{k=0}^{K} \tilde{P}_{ik}[\varphi(s)]^k[1-(\frac{3}{4}-s/2)k/N]^l = \sum_{k=0}^{K} \tilde{P}_{ik}[\varphi(s)]^k + O(1/N)$ and then $= \sum_{k=0}^{N} \tilde{P}_{ik}[\varphi(s)]^k + O(1/N) - 2\varepsilon = [\varphi_2(s)]^i + O(1/N) - 2\varepsilon$. The result of (49) now routinely follows. An analogous argument proves

$$\lim_{N \to \infty} \sum_{j=0}^{N} \tilde{P}_{ij}^{(r)}(N) s^j \geqslant [\varphi_r(s)]^i \tag{50}$$

uniformly for $0 \leqslant s \leqslant 1$. The conjunction of (48) and (50) establishes

$$\lim_{N \to \infty} \sum_{j=0}^{N} \tilde{P}_{ij}^{(r)}(N) s^j = [\varphi_r(s)]^i \tag{51}$$

uniformly in $0 \leqslant s \leqslant 1$ for each fixed i, $0 \leqslant i \leqslant N$, and integer $r = 1,2,3,\ldots$.

Let $T_0^{(N)}$ ($T_D^{(N)}$) be the random variable of the time to loss (detection) of allele a for the process (18), and T_0^{*} (T_D^{*}) that for the corresponding branching process with killing. Obviously $P\{T_0^{(N)} \leqslant r \mid X^{(N)}(0) = i\} = P_{i0}^{(r)}(N)$, which converges to $[\varphi_r(0)]^i = P_{i0}^{*(r)}$. Obviously for each integer t, $P\{T_D^{(N)} > t \mid X_N(0) = i\} = \sum_{j=0}^{N} \tilde{P}_{ij}^{(t)}(N)$ converges to $P\{T_D^{*} > t \mid X(0) = i\} = [\varphi_t(1)]^i$. Actually, the convergence is very strong, such that all moments of $T_D^{(N)}$ and $T_0^{(N)}$ can be well approximated by those of T_D^{*} and T_0^{*}, respectively.

The same kind of analysis applies, *mutatis mutandis*, to all the discrete processes and their branching-process limits of Sections 2 and 3.

REFERENCES

1 K. Athreya and P. Ney, *Branching Processes*, Springer, New York, 1972.

2 M. S. Bartlett and J. B. S. Haldane, The theory of inbreeding with forced heterozygosis, *J. Genet.* 31:327–340 (1935).

3 W. J. Ewens, *Mathematical Population Genetics*, Springer, New York, 1979.

4 J. W. James, The time of detection of sex-linked recessives in small populations, *Genet. Res. (Camb.)* 34:11–17 (1979).

5 S. Karlin and S. Tavaré, The detection of a recessive visible gene in finite populations, *Genet. Res. (Camb.)* 37:33–46 (1981).

6 S. Karlin and S. Tavaré, The detection of particular genotypes in finite populations. I. Natural selection effects, *Theoret. Population Biol.* 19:187–214 (1981).

7 S. Karlin and S. Tavaré, The detection of particular genotypes in finite populations. II. The effects of partial penetrance and family structure, *Theoret. Population Biol.* 19:215–229 (1981).

8 S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, 2nd ed., Vol. I, Academic Press, New York, 1975.

9 S. Karlin, Equilibrium behavior of population genetic models with non-random mating, J Appl. Prob 5,231–313 and 487–566.

10 M. Kimura, The length of time required for a selectively neutral mutant to reach fixation through random drift in a finite population, *Genet. Res.* (*Camb.*) 15:131–133 (1980).

11 A. Robertson, The time to detection of recessive visible genes in small populations, *Genet. Res.* (*Camb.*) 31:255–264 (1978).

12 G. A. Watterson, Nonrandom mating, and its effect on the rate of approach to homozygosity, *Ann. Human Genetics* 23:204–220 (1959).