Journal of Mathematical Biology 1, 317—320 (1975)

The Genetic Balance between Random Sampling and Random Population Size

C. C. Heyde and **E. Seneta**, Canberra

Received October 28, 1974

Summary

This note pertains to a generalized model for random fluctuation of allele frequency, where the population size is permitted to fluctuate randomly from generation to generation. Martingale methods are applied to discuss in two propositions, respectively, necessary and sufficient conditions for $P(Y(1-Y)>0)>0$, where Y is the (almost sure) limiting frequency of one allele. Under an additional restriction a necessary and sufficient condition results. Some simple remarks are made on further description of the distribution of Y.

1. Introduction

We concern ourselves with the standard problem of a single diploid diallelic autosomal locus (genotypes AA, Aa, aa) and non-overlapping generations, in the absence of external pressures, in a random-mating population whose size remains finite, and is denoted by $N_n (\geq 1)$ at the n-th generation, $n \geq 0$. In an extension of a classical fixed size model of S. Wright, one of the authors (Seneta 1974) assumed that the population size may vary, even randomly.

Specifically, the model states:

(a) the process $\{X_n, N_n\}$, $n \geq 0$, is bivariate Markov, X_n denoting the number of a alleles in the n-th generation; and

(b) the distribution of X_{n+1}, given N_{n+1}, N_n, X_n is binomial, and given by

$$\binom{2N_{n+1}}{j} Y_n^j (1-Y_n)^{2N_{n+1}-j}, \quad 0 \leq j \leq 2N_{n+1}, \tag{1}$$

where $Y_n = X_n/(2N_n)$ is the a-allele proportion of alleles in the n-th generation.

Write $\mathscr{F}_n$ for the σ-field generated by $\{(X_0, N_0), (X_1, N_1), \ldots, (X_n, N_n)\}$, $n \geq 0$. It was observed in the cited paper that $\{Y_n, \mathscr{F}_n, n \geq 0\}$ is a martingale, and so $Y_n \to Y$ a. s. as $n \to \infty$ for some random variable Y with $0 \leq Y \leq 1$, and $EY = EY_n$, $n \geq 0$.

One of the main issues of interest is, then: when is the limit distribution trivial i.e. $P(\{Y=0\} \cup \{Y=1\}) = 1$, or equivalently $P(Y(1-Y)=0)=1$, so fixation is certain; and when not, so that a situation of balanced polymorphism occurs with positive probability.

In the case where $\{N_n\}$ is purely deterministically varying, it was noted (Seneta, 1974) that a necessary and sufficient condition for non-triviality is $\sum_{n=0}^{\infty} 1/N_{n+1} < \infty$;

DOI 10.1007/978-1-4419-5823-5_39, © Springer Science+Business Media, LLC 2010

providing as always the process does not begin in a state of fixation. The more general situation where the sequence $\{N_n\}$ is permitted to be random, in relation to the same question, is the subject of the present note; which is also, again, an illustration of the applicability of martingale methods to the present model. One direct deduction from the result on the deterministic case to a stochastic one in relation to $\{N_n\}$ may be worth mentioning at this point; if $\sum 1/N_{n+1} < \infty$ a.s. on a set A of histories of positive probability, and for almost every given sequence $\{N_n\}$ in the set A, the sequence $\{X_n\}$ is still Markovian, then the limit distribution is not trivial. (This may most easily be seen by postulating the contrary i.e. triviality). Results related to this one are given in Theorem 2 below, which gives other sufficient conditions for non-triviality.

The starting point of our deductions is the relation

$$E(Y_{n+1}(1-Y_{n+1}) \mid \mathscr{F}_n) = (1 - E\{(2N_{n+1})^{-1} \mid \mathscr{F}_n\})\, Y_n(1-Y_n). \tag{2}$$

This follows from the above basic assumptions (a) and (b), since these imply via (1) that

$$E(Y_{n+1} \mid N_{n+1}, \mathscr{F}_n) = Y_n, \ \mathrm{Var}(Y_{n+1} \mid N_{n+1}, \mathscr{F}_n) = (2N_{n+1})^{-1}\, Y_n(1-Y_n)$$

whence

$$E(Y_{n+1}(1-Y_{n+1}) \mid N_{n+1}, \mathscr{F}_n) = (1-(2N_{n+1})^{-1})\, Y_n(1-Y_n).$$

From (2) we note immediately that $E(Y_n(1-Y_n)) \downarrow$ as n increases (and, indeed, that $\{Y_n(1-Y_n), \mathscr{F}_n, n \geq 0\}$ is a supermartingale).

2. Main Results

Theorem 1: *If* $\sum_{n=0}^{\infty} E(N_{n+1}^{-1} \mid \mathscr{F}_n) = \infty$ *a.s. on some set A of histories of the process* $\left(\textit{or equivalently } \sum_{n=0}^{\infty} N_{n+1}^{-1} = \infty \textit{ a.s. on } A\right)$ *then fixation occurs on A (i.e. $Y(1-Y)=0$ a.s. on A).*

Proof: Define the martingale $\{U_n, \mathscr{F}_n, n \geq 1\}$ by

$$U_n = \sum_{k=1}^{n} [Y_k(1-Y_k) - E(Y_k(1-Y_k) \mid \mathscr{F}_{k-1})], \ n \geq 1.$$

Using (2), we have

$$\begin{aligned} U_n &= \sum_{k=1}^{n} Y_k(1-Y_k) - \sum_{k=0}^{n-1} [1 - E((2N_{k+1})^{-1} \mid \mathscr{F}_k)]\, Y_k(1-Y_k) \\ &= Y_n(1-Y_n) - Y_0(1-Y_0) + \sum_{k=0}^{n-1} E((2N_{k+1})^{-1} \mid \mathscr{F}_k)\, Y_k(1-Y_k). \end{aligned}$$

$\{U_n, \mathscr{F}_n, n \geq 1\}$ is a zero-mean martingale with bounded increments and so, using Proposition IV-6-3 (of Neveu, 1965) $\liminf_{n\to\infty} U_n = -\infty$ a.s., $\limsup_{n\to\infty} U_n = +\infty$ a.s. on the set where U_n does not converge a.s. It follows that U_n must converge a.s. as $n \to \infty$ since $U_n \geq -Y_0(1-Y_0)$ for all n. Since $Y_n \to Y$ a.s., we have $Y_n(1-Y_n) \to Y(1-Y)$ a.s. and it follows that

$$\sum_{n=0}^{\infty} E(N_{n+1}^{-1} \mid \mathscr{F}_n)\, Y_n(1-Y_n) < \infty \text{ a.s.} \tag{3}$$

It is clear from (3) that if $\sum_{n=0}^{\infty} E(N_{n+1}^{-1} \mid \mathscr{F}_n) = \infty$ a.s. on A then $Y_n(1-Y_n) \to 0$ on A, i.e. fixation occurs on A. That $\sum_{n=0}^{\infty} N_{n+1}^{-1} = \infty$ a.s. on A if and only if $\sum_{n=0}^{\infty} E(N_{n+1}^{-1} \mid \mathscr{F}_n) = \infty$ a.s. on A follows from another application of Proposition IV-6-3 of Neveu. This completes the proof.

It is worth noting that in complementary form the result of Theorem 1 says that we must have $\sum_{n=0}^{\infty} E(N_{n+1}^{-1} \mid \mathscr{F}_n) < \infty$ a.s. on the set $\{Y(1-Y) > 0\}$ (and equivalently $\sum_{n=0}^{\infty} N_{n+1}^{-1} < \infty$ a.s. on $\{Y(1-Y) > 0\}$).

It has not been possible to obtain a converse to Theorem 1. The following partial result does however, cover many cases of interest.

Theorem 2: *If $E(N_{n+1}^{-1} \mid \mathscr{F}_n) \le \alpha_n$ a.s. for all n where $\{\alpha_n\}$ is a sequence of positive constants with $\sum \alpha_n < \infty$, or if $N_n \ge \beta_n$ a.s. for all n where $\{\beta_n\}$ is a sequence of positive constants with $\sum \beta_n^{-1} < \infty$, then $P(Y(1-Y) > 0) > 0$.*

Proof: We use the martingale $\{U_n, \mathscr{F}_n, n \ge 1\}$ introduced in the proof of Theorem 1. First note that, under the conditions of the theorem,

$$E \sum_{k=0}^{\infty} E((2N_{k+1})^{-1} \mid \mathscr{F}_k)\, Y_k(1-Y_k) < \infty.$$

Then, for any fixed $n \ge 0$ and $N > n$ we have from (2) that

$$\begin{aligned} E \sum_{k=n}^{N} Y_k(1-Y_k)\, E((2N_{k+1})^{-1} \mid \mathscr{F}_k) \\ &= E[U_{N+1} - U_n - Y_{N+1}(1-Y_{N+1}) + Y_n(1-Y_n)] \\ &= -E\, Y_{N+1}(1-Y_{N+1}) + E\, Y_n(1-Y_n) \\ &\to -E\, Y(1-Y) + E\, Y_n(1-Y_n) \end{aligned}$$

as $N \to \infty$ by dominated convergence since $Y_{N+1}(1-Y_{N+1}) \to Y(1-Y)$ a.s. as $N \to \infty$. The monotone convergence theorem then gives

$$E \sum_{k=n}^{\infty} Y_k(1-Y_k)\, E((2N_{k+1})^{-1} \mid \mathscr{F}_k) = -E\, Y(1-Y) + E\, Y_n(1-Y_n) \tag{4}$$

and hence

$$E \sum_{k=n}^{\infty} Y_k(1-Y_k)\, E((2N_{k+1})^{-1} \mid \mathscr{F}_k) \le E\, Y_n(1-Y_n)$$

with equality if and only if $E\, Y(1-Y) = 0$ (i.e. $P(Y(1-Y) > 0) = 0$).

Now suppose that $E((2N_{k+1})^{-1} \mid \mathscr{F}_k) \le \alpha_k$ a.s. for all k where $\{\alpha_k\}$ is a sequence of constants with $\sum \alpha_k < \infty$. We can choose n so large that $\sum_{k=n}^{\infty} \alpha_k < 1$ and then

$$E \sum_{k=n}^{\infty} Y_k (1 - Y_k) E\left((2 N_{k+1})^{-1} \mid \mathscr{F}_k\right)$$

$$\leq \sum_{k=n}^{\infty} E\left(Y_k (1 - Y_k)\right) \alpha_k$$

$$\leq E\left(Y_n (1 - Y_n)\right) \sum_{k=n}^{\infty} \alpha_k < E\, Y_n (1 - Y_n)$$

since we have noted $E\left(Y_k (1 - Y_k)\right) \downarrow$ as k increases. This assures, via (4), that $E\left(Y(1 - Y)\right) > 0$, i.e. $P\left(Y(1 - Y) > 0\right) > 0$. Similar reasoning applies if $N_n > \beta_n$ a. s. for all n where $\{\beta_n\}$ is a sequence of positive constants with $\sum \beta_n^{-1} < \infty$. This completes the proof.

Two corollaries of Theorems 1 and 2 may be worth mentioning. First, if $E(N_{n+1}^{-1} \mid \mathscr{F}_n) = E(N_{n+1}^{-1})$, then $P\left(Y(1 - Y) > 0\right) > 0$ if and only if $\sum E(N_{n+1}^{-1}) < \infty$, a direct extension of the purely deterministic situation of $\{N_n\}$, which can also be deduced directly from (2). Secondly, if N_{n+1} is totally independent of $\mathscr{F}_n$, $n \geq 0$, i. e. of the past history of the bivariate process, which covers the situation of so-called "randomly fluctuating environments", then clearly this same necessary and sufficient condition applies.

3. The Limit Distribution

By noting that for $n \geq 0$, and y fixed at 0 or 1,

$$\{Y_n = y\} \Rightarrow \{Y_{n+1} = y\} \Rightarrow \{Y = y\}$$

we obtain

$$P(Y_n = y) \uparrow \alpha(y) \leq P(Y = y), \tag{5}$$

so for any given n, $P(Y_n = y)$ may serve as a lower bound for $P(Y = y)$. Consequently, also $1 - \alpha(0) - \alpha(1) \geq P(0 < Y < 1)$. That strict inequality will sometimes apply in (5) may be expected from other contexts. One also has available the inequalities valid for any random variable Y on $[0, 1]$, whose mean is $c: P(Y = 1) \leq c$, $P(|Y - c| \geq t) \leq t^{-2} \max\{c^2, (1 - c)^2\}$, these being basically Markov's and Chebyshev's inequalities, respectively. In our case, $c = E\,Y = E\,Y_0$ is generally known.

References

Neveu, J.: Mathematical Foundations of the Calculus of Probability. San Francisco: Holden-Day 1965.

Seneta, E.: A note on the balance between random sampling and population size. (On the 30th anniversary of G. Malécot's paper.) Genetics 77, 607—610 (1974).

Dr. C. C. Heyde
Dr. E. Seneta
Department of Statistics, S.G.S.
Australian National University
P.O. Box 4
Canberra, A.C.T. 2600
Australia

Printed in Austria
Druck: Paul Gerin, A-1021 Wien