CHANGE OF GENOTYPE FREQUENCIES UNDER SELECTION IN A SELF-POLLINATED POPULATION

ERIC RENSHAW

Department of Statistics, University of Edinburgh, Edinburgh, U.K.

SUMMARY

Whilst genotype frequencies under full selfing may be determined explicitly for a self-pollinated population in which genotypes differ in fitness, the non-linearity of the mixed selfing and random outcrossing model necessitates the development of an asymptotic approach.

1. INTRODUCTION

THE general theory of the change of genotype frequencies under selfing when genotypes differ in fitness, including the extension to selfing-with-crossing and random outbreeding, is described by Workman and Allard (1962). Applications to experimental populations were initiated by Allard and his colleagues, namely: Jain and Allard (1960); Allard and Jain (1962); Allard and Workman (1963). We shall first consider closed expressions for the nth generation frequencies under full selfing, and then develop a general asymptotic technique within the non-linear context of mixed selfing and random outcrossing.

2. EXPRESSIONS WITH FULL SELFING

Let $f_1^{(n)}$, $f_2^{(n)}$ and $f_3^{(n)}$ denote the relative frequencies in the nth generation of genotypes AA, Aa and aa at a single locus. Suppose also that the fitness of the AA, Aa and aa genotypes is 1, s_2 and s_3, respectively. Then the recurrence equations relating the genotype frequencies in generations n and $n+1$, assuming full selfing, are given by

$$f_1^{(n+1)} = (f_1^{(n)} + \tfrac{1}{4}s_2 f_2^{(n)})/W^{(n)} \tag{1}$$

$$f_2^{(n+1)} = (\tfrac{1}{2}s_2 f_2^{(n)})/W^{(n)} \tag{2}$$

$$f_3^{(n+1)} = (s_3 f_3^{(n)} + \tfrac{1}{4}s_2 f_2^{(n)})/W^{(n)} \tag{3}$$

where $n = 1, 2, 3, \ldots$,

$$W^{(n)} = f_1^{(n)} + s_2 f_2^{(n)} + s_3 f_3^{(n)}$$

and

$$f_1^{(n)} + f_2^{(n)} + f_3^{(n)} = 1.$$

Similar equations have been discussed with varying degrees of completeness by Haldane (1956), Hayman and Mather (1956) and Reeve (1957), solutions being derived via either difference equation or matrix techniques. Karlin (1968) provides a comprehensive account of various mathematical approaches.

On disregarding the scaling factor $W^{(n)}$ relations (1)–(3) are linear in the $f_i^{(n)}$ ($i = 1, 2, 3$), so let us first consider the three equations

$$p_{n+1} = p_n + \tfrac{1}{4}s_2 q_n \tag{4}$$

$$q_{n+1} = \tfrac{1}{2}s_2 q_n \tag{5}$$

$$r_{n+1} = s_3 r_n + \tfrac{1}{4}s_2 q_n. \tag{6}$$

Their solution follows directly either from a standard matrix approach (Nagylaki, 1977, pp. 96–99), or from noting that equation (5) immediately yields

$$q_n = q(\tfrac{1}{2}s_2)^n \tag{7}$$

whence substitution of (7) into (4) and (6) gives standard single-variable difference equations (see, for example, Stephenson (1961)) with the solutions

$$p_n = p + [\tfrac{1}{2}qs_2/(s_2 - 2)][(\tfrac{1}{2}s_2)^n - 1] \tag{8}$$

$$r_n = rs_3^n + [\tfrac{1}{2}qs_2/(s_2 - 2s_3)][(\tfrac{1}{2}s_2)^n - s_3^n] \tag{9}$$

subject to the initial conditions $f_1^{(0)} = p$, $f_2^{(0)} = q$ and $f_3^{(0)} = r$.

We may now form $f_1^{(n)} = p_n/W^{(n)}$, $f_2^{(n)} = q_n/W^{(n)}$ and $f_3^{(n)} = r_n/W^{(n)}$ by dividing (8), (7) and (9) by

$$
\begin{aligned}
W^{(n)} = p_n + q_n + r_n = {}&\{p - (\tfrac{1}{2}qs_2)(s_2 - 2)^{-1}\} \\
&+ q(\tfrac{1}{2}s_2)^n\{1 + (\tfrac{1}{2}s_2)[(s_2 - 2)^{-1} + (s_2 - 2s_3)^{-1}]\} \\
&+ s_3^n\{r - (\tfrac{1}{2}qs_2)(s_2 - 2s_3)^{-1}\}
\end{aligned}
$$

to give

$$
\begin{aligned}
f_1^{(n)} &= \{p + [\tfrac{1}{2}qs_2/(s_2 - 2)][(\tfrac{1}{2}s_2)^n - 1]\}/W^{(n)} \\
f_2^{(n)} &= \{q(\tfrac{1}{2}s_2)^n\}/W^{(n)} \\
f_3^{(n)} &= \{rs_3^n + [\tfrac{1}{2}qs_2/(s_2 - 2s_3)][(\tfrac{1}{2}s_2)^n - s_3^n]\}/W^{(n)}.
\end{aligned} \tag{10}
$$

Of particular interest is the behaviour of these relative frequencies as n becomes large. If $0 \le \tfrac{1}{2}s_2, s_3 < 1$ then $\lim_{n \to \infty} f_1^{(n)} = 1$ whilst both $f_2^{(n)}$ and $f_3^{(n)}$ decrease (almost geometrically) to zero but not necessarily at the same rate. For

$$
f_3^{(n)}/f_2^{(n)} \sim
\begin{cases}
(\tfrac{1}{2}s_2)/(s_2 - 2s_3) & s_2 > 2s_3 \\
(r/q + \tfrac{1}{2}n) & s_2 = 2s_3 \\
(2s_3/s_2)^n[(r/q) - (\tfrac{1}{2}s_2)/(s_2 - 2s_3)] & s_2 < 2s_3.
\end{cases} \tag{11}
$$

Different limiting relations apply with other values of s_2 and s_3. For example, if s_3 exceeds both 1 and $\tfrac{1}{2}s_2$ then both $f_1^{(n)}$ and $f_2^{(n)}$ approach zero and $f_3^{(n)}$ approaches 1, whilst if $s_2 > 2$ and $s_3 < \tfrac{1}{2}s_2$ then in the limit as $n \to \infty$ $f_1^{(n)} : f_2^{(n)} : f_3^{(n)} = (\tfrac{1}{2}s_2)/(s_2 - 2) : 1 : (\tfrac{1}{2}s_2)/(s_2 - 2s_3)$. The condition for $\lim_{n \to \infty} f_2^{(n)} > 0$ is slightly more general than this latter one, being $s_2 \ge 2$, $s_3 \le \tfrac{1}{2}s_2$.

3. ASYMPTOTIC RESULTS FOR MIXED SELFING AND RANDOM OUTCROSSING

The above approach relies on the linearity of equations (4)–(6), so in order to investigate non-linearity let us allow random outcrossing with probability t and selfing with probability $z = 1 - t$. The recurrence relations (1)–(3) are now replaced by the relations

$$G_1^{(n)} = t(f_1^{(n)} + \tfrac{1}{2}s_2 f_2^{(n)})^2 + z(f_1^{(n)} + \tfrac{1}{4}s_2 f_2^{(n)}) \tag{12}$$

$$G_2^{(n)} = 2t(f_1^{(n)} + \tfrac{1}{2}s_2 f_2^{(n)})(s_3 f_3^{(n)} + \tfrac{1}{2}s_2 f_2^{(n)}) + z(\tfrac{1}{2}s_2 f_2^{(n)}) \tag{13}$$

$$G_3^{(n)} = t(s_3 f_3^{(n)} + \tfrac{1}{2}s_2 f_2^{(n)})^2 + z(s_3 f_3^{(n)} + \tfrac{1}{4}s_2 f_2^{(n)}) \tag{14}$$

with

$$f_i^{(n+1)} = G_i^{(n)} / W^{(n)} \quad \text{and} \quad W^{(n)} = G_1^{(n)} + G_2^{(n)} + G_3^{(n)}. \tag{15}$$

Conditions (15) ensure that $f_1^{(n)} + f_2^{(n)} + f_3^{(n)} = 1$.

If $s_2 = s_3 = 1$ then, from the Hardy–Weinberg rule, the quadratic terms in (12)–(14) reduce to a matrix of constants and Workman and Allard (1962) exploited the resulting linearity of equations (12)–(14) to derive a matrix solution for the relative frequencies $f_i^{(n)}$, an equilibrium solution being given by Hayman (1953). However, if $s_2, s_3 \neq 1$ then this simplification no longer applies and an alternative approach is necessary.

First we examine equilibrium solutions by letting $f_i^{(n)} \to f_i$ as $n \to \infty$. As $f_1 + f_2 + f_3 = 1$ we have

$$W^{(n)} \to t(f_1 + s_2 f_2 + s_3 f_3)^2 + z(f_1 + s_2 f_2 + s_3 f_3),$$

and so in the limit equations (12)–(15) become

$$f_1[t(f_1 + s_2 f_2 + s_3 f_3)^2 + z(f_1 + s_2 f_2 + s_3 f_3)]$$
$$= t(f_1 + \tfrac{1}{2}s_2 f_2)^2 + z(f_1 + \tfrac{1}{4}s_2 f_2) \tag{16}$$

$$f_2[t(f_1 + s_2 f_2 + s_3 f_3)^2 + z(f_1 + s_2 f_2 + s_3 f_3)]$$
$$= 2t(f_1 + \tfrac{1}{2}s_2 f_2)(s_3 f_3 + \tfrac{1}{2}s_2 f_2) + z(\tfrac{1}{2}s_2 f_2) \tag{17}$$

$$f_3[t(f_1 + s_2 f_2 + s_3 f_3)^2 + z(f_1 + s_2 f_2 + s_3 f_3)]$$
$$= t(s_3 f_3 + \tfrac{1}{2}s_2 f_2)^2 + z(s_3 f_3 + \tfrac{1}{4}s_2 f_2). \tag{18}$$

On replacing f_1 by $1 - f_2 - f_3$ in expressions (17) and (18) we see that f_2 and f_3 are defined by the intersection of two cubic surfaces, and so in general may take several values. For admissible solutions we require f_2, $f_3 > 0$; $f_2 + f_3 \leqq 1$ and direct inspection reveals $f_1 = 1$, $f_2 = f_3 = 0$ and $f_3 = 1$, $f_1 = f_2 = 0$ as being two of them. For the remainder of this paper we shall restrict our attention to the former solution with $0 < s_2, s_3 < 1$.

Let us therefore write $f_1^{(n)} = 1 + o(1)$ and $f_2^{(n)}, f_3^{(n)} = o(1)$, so that $W^{(n)} = 1 + o(1)$ where $o(1)$ denotes terms which approach zero as n tends to infinity. Denote

$$\alpha_n = f_2^{(n)} / f_3^{(n)}.$$

Then as

$$\frac{f_2^{(n+1)}}{f_3^{(n+1)}} = \frac{G_2^{(n)} / W^{(n)}}{G_3^{(n)} / W^{(n)}} \sim \frac{2t(s_3 f_3^{(n)} + \tfrac{1}{2}s_2 f_2^{(n)}) + z(\tfrac{1}{2}s_2 f_2^{(n)})}{z(s_3 f_3^{(n)} + \tfrac{1}{4}s_2 f_2^{(n)})}$$

we have

$$\alpha_{n+1} \sim [2t(s_3 + \tfrac{1}{2}s_2\alpha_n) + z(\tfrac{1}{2}s_2\alpha_n)]/[z(s_3 + \tfrac{1}{4}s_2\alpha_n)].$$

Thus if $\alpha_n \to \alpha$ as $n \to \infty$, then

$$(\tfrac{1}{4}zs_2)\alpha^2 + (zs_3 - ts_2 - \tfrac{1}{2}zs_2)\alpha - 2ts_3 = 0$$

which has the one admissible root

$$\alpha = (\tfrac{1}{2}zs_2)^{-1}[-(zs_3 - ts_2 - \tfrac{1}{2}zs_2) + \sqrt{\{(zs_3 - ts_2 - \tfrac{1}{2}zs_2)^2 + 2tzs_2s_3\}}]. \tag{19}$$

Note that if $t = 0$ and $z = 1$ then the situation reduces to that of section 2, with (19) simplifying to $\alpha = \max\{0, (s_2 - 2s_3)/(\tfrac{1}{2}s_2)\}$ in agreement with the limiting form of (11) as $n \to \infty$.

Next we investigate the behaviour of $f_2^{(n)}$ and $f_3^{(n)}$ separately by writing

$$\beta_n = f_3^{(n+1)}/f_3^{(n)}.$$

From (14) and (15)

$$\frac{f_3^{(n+1)}}{f_3^{(n)}} = \frac{G_3^{(n)}}{f_3^{(n)}[1 + o(1)]} \sim \frac{z(s_3 f_3^{(n)} + \tfrac{1}{4}s_2 f_2^{(n)})}{f_3^{(n)}}$$

and so

$$\beta_n \sim z(s_3 + \tfrac{1}{4}s_2\alpha_n).$$

Thus if $\beta_n \to \beta$ as $n \to \infty$, then

$$\beta = z(s_3 + \tfrac{1}{4}s_2\alpha). \tag{20}$$

On combining these results we have

$$f_3^{(n)} \sim k\beta^n \quad \text{and} \quad f_2^{(n)} \sim k\alpha\beta^n \tag{21}$$

for some constant $k > 0$ which depends on $f_2^{(0)}$ and $f_3^{(0)}$. Note that (21) implies

$$f_1^{(n)} = 1 - k(1 + \alpha)\beta^n \tag{22}$$

which may be shown to satisfy (12).

To illustrate the way in which α and β change with z and t ($0 \leq z = 1 - t \leq 1$), their behaviour is shown in fig. 1 for $s_2 = 0.9$ and $s_3 = 0.3$. Whilst β is well-behaved over the entire region $0 \leq z \leq 1$, α is well-behaved only for values of z away from zero. In this example, $\log \alpha$ is almost linear in the range $0.30 \leq z \leq 0.85$ but approaches infinity as z nears zero.

As expressions (21) and (22) denote probabilities we clearly require both $\alpha \geq 0$ and $0 \leq \beta < 1$. The first follows from (19), whence (20) gives $\beta = zs_3 + \tfrac{1}{4}zs_2\alpha \geq 0$. Thus we need only consider $\beta < 1$, i.e.,

$$zs_3 + \tfrac{1}{4}zs_2\alpha < 1$$

which from (19) becomes

$$\sqrt{\{(zs_3 - ts_2 - \tfrac{1}{2}zs_2)^2 + 2tzs_2s_3\}} < 2 - zs_3 - ts_2 - \tfrac{1}{2}zs_2. \tag{23}$$

On squaring both sides this reduces to

$$z(s_2 - 2s_3 + s_2s_3) > 2(s_2 - 1). \tag{24}$$

Not only is this inequality satisfied if $s_2, s_3 \leq 1$, but also the right-hand side of (23) is non-negative, as required. For such non-negativity is implicitly
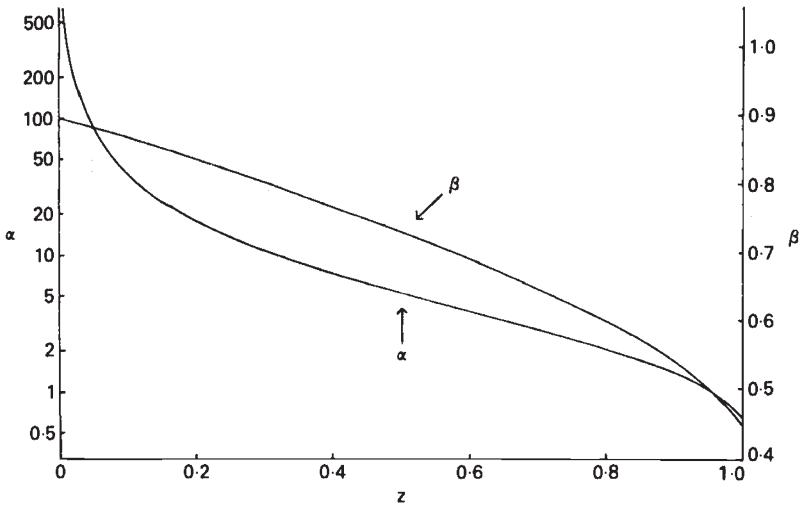
FIG. 1.—Change in α (logarithmic scale) and β with z, evaluated from expressions (19) and (20), respectively, for $s_2 = 0\cdot9$, $s_3 = 0\cdot3$.

assumed in the act of squaring, and so for general s_2, s_3 the requirement $\beta < 1$ necessitates the condition

$$2 - zs_3 - ts_2 - \tfrac{1}{2}zs_2 \geqq 0$$

in addition to (24). After a little manipulation we may write these conditions in the form, "for solutions (21) and (22) to hold, we require $z(s_2 - 2s_3)$ to exceed both $2(s_2 - 2)$ and $2(s_2 - 1) - zs_2s_3$".

Expressions (19) and (20) for α and β may be further manipulated to yield more detailed information in several special cases. For example, in the case of equal fitness ($s_2 = s_3 = s$) then $\alpha = 4t/z$ and $\beta = s$, the former being independent of s and the latter being independent of z. Whilst if s_2 and s_3 are only just less than 1, so that $s_i = 1 - \varepsilon_i$ ($i = 2, 3$) for $0 < \varepsilon_i \ll 1$, then $\beta \sim 1 - (2t\varepsilon_2 + z\varepsilon_3)/(1 + t)$. The infinite behaviour of α as $z \to 0$, illustrated in fig. 1, also follows with:

$$\alpha \sim (4/z) - (2/s_2)(s_2 + s_3) \quad \text{and} \quad \beta \sim s_2 - \tfrac{1}{2}z(s_2 - s_3) \quad \text{as} \quad z \to 0;$$

$$\alpha \to \max\{0, (s_2 - 2s_3)/(\tfrac{1}{2}s_2)\} \quad \text{and} \quad \beta \to \max\{\tfrac{1}{2}s_2, s_3\} \quad \text{as} \quad z \to 1.$$

4. CONCLUSIONS

Whilst the approach of section 2 relies on the ability to express the recurrence equations for the relative frequencies in linear form, in section 3 we develop a general asymptotic technique to enable the analysis of non-linear equations. Although the exact determination of relative frequencies for specific known parameter values still necessitates numerical computation via the recursion formulae (12)–(15), the asymptotic arguments described throw considerable light on how the fitness parameters s_2, s_3 and mixing coefficients z, t affect the development of the $\{f_i^{(n)}\}$. Of particular value is the condition on s_2, s_3 and z necessary for convergence to $(1, 0, 0)$—

conditions for convergence to other admissible equilibrium solutions may be developed in a similar manner. However, the importance of this asymptotic approach lies not so much in its success with this particular model, but that the techniques employed are of perfectly general application and should yield results in many other interesting non-linear situations.

5. REFERENCES

ALLARD, R. W., AND JAIN, S. K. 1962. Population studies in predominantly self-pollinated species. II. Analysis of quantitative genetic changes in a bulk-hybrid population of barley. *Evolution, 16,* 90–101.

ALLARD, R. W., AND WORKMAN, P. L. 1963. Population studies in predominantly self-pollinated species. IV. Seasonal fluctuations in estimated values of genetic parameters in lima bean populations. *Evolution, 17,* 470–480.

HALDANE, J. B. S. 1956. The conflict between inbreeding and selection. I. Self-fertilization. *J. Genet., 54,* 56–63.

HAYMAN, B. I. 1953. Mixed selfing and random mating when homozygotes are at a disadvantage. *Heredity, 7,* 185–192.

HAYMAN, B. I., AND MATHER, K. 1956. Inbreeding when homozygotes are at a disadvantage: a reply. *Heredity, 10,* 271–274.

JAIN, S. K., AND ALLARD, R. W. 1960. Population studies in predominantly self-pollinated species. I. Evidence for heterozygote advantage in a closed population of barley. *Proc. Nat. Acad. Sci., 46,* 1371–1377.

KARLIN, S. 1968. Equilibrium behaviour of population genetic models with non-random mating. *J. Appl. Prob., 5,* 231–313.

NAGYLAKI, T. 1977. *Selection in One- and Two-Locus Systems.* Springer-Verlag, Berlin.

REEVE, E. C. R. 1957. Inbreeding with selection and linkage. I. Selfing. *Ann. Hum. Genet., 21,* 277–288.

STEPHENSON, G. 1961. *Mathematical Methods for Science Students.* Longman, London.

WORKMAN, P. L., AND ALLARD, R. W. 1962. Population studies in predominantly self-pollinated species. III. A matrix model for mixed selfing and random outcrossing. *Proc. Nat. Acad. Sci., 48,* 1318–1325.