THE TIME REQUIRED FOR ALLELE FREQUENCY CHANGE

TAI-HE FAN[✉1], SHUHAO SUN[2] and PING-AN HE[1]

Abstract

In evolutionary theory, a key issue in selection theory is the expected time for a given amount of allele frequency change to occur. Crow and Kimura, by assuming weak selection, presented explicit results for several important cases of the directional selection and of the stochastic process. Those results played an important role in the theory of population genetics. In this paper, first we show that the weak selection assumption can be removed from most of the results of Crow and Kimura, and then we generalize those results to the most general selection model. Next, we estimate the errors of the deterministic formulae produced by proving that the deterministic formulae are limits of the corresponding stochastic formulae when the size of the population tends to infinity. Finally, we present a result which removes the restriction of Kimura's corresponding results for a favourite recessive selection model, and we also observe that the conclusion made by Kimura about the favourite dominant selection might not be correct.

1. Introduction

A key issue in selection theory is the expected time for a given amount of allele frequency change to occur. Crow and Kimura presented explicit results for several important cases of the directional selection (or purifying selection) [1, (5.13)–(5.14) and (5.3.15)]; the values of fitnesses were $1 + s, 1 + hs$ and 1 with s as the selective coefficient. In the following discussion, the variables x and y are the frequencies of A and a alleles, respectively, with $x + y = 1$ and t being the time required for allele A moving from x to x_i.

---

[1]Faculty of Sciences, Zhejiang Sci-Tech University, Hangzhou, 310018, China;
e-mail: taihefan@163.com, pinganhe@zstu.edu.cn.
[2]School of Mathematical Science, Monash University, Australia; e-mail: shuhaosun888@hotmail.com.

If the fitnesses are additive ($h = 1/2$), then, by assuming $s(1 - x) \approx 0$,

$$t = \frac{2}{s} \ln \frac{(1 - x)x_t}{(1 - x_t)x}, \tag{1.1}$$

whereas the allele a is completely recessive to A ($h = 1$), by assuming $s(1 - x)^2 \approx 0$,

$$t = \frac{1}{s}\left[\frac{1}{1 - x_t} - \frac{1}{1 - x} + \ln \frac{(1 - x)x_t}{(1 - x_t)x} \right]. \tag{1.2}$$

If the allele a is completely dominant to A ($h = 0$), then, by assuming $s(1 - x^2) \approx 0$,

$$t = \frac{1}{s}\left[\ln \frac{x_t(1 - x)}{x(1 - x_t)} + \frac{1}{x} - \frac{1}{x_t} \right]. \tag{1.3}$$

These results play an important role in theoretical population genetics and have not been improved yet (see, for example, the article by Hartl and Clark [4, (6.15)–(6.17)] and the books by Hedrick [5, (3.6b)–(3.6c)], Walsh and Lynch [8, (7.3c)–(7.3e)] and Ewens [2, (1.28)]).

For Kimura's stochastic process, consider a continuous random variable x_t indexed by continuous time t. If $\delta_t = x_{t+\delta} - x_t$ satisfy

$$E(\delta_x | x_t = x) = m(x)\delta_t + o(\delta_t),$$
$$\sigma^2(\delta_x | x_t = x) = v(x)\delta_t + o(\delta_t),$$
$$E(|\delta_t|^k) = o(\delta_t)$$

for $k \geq 3$, the random variable x_t is said to be a *diffusion process*. The infinitesimal mean $m(x)$ and infinitesimal variance $v(x)$ correspond to the mean and variance of the process over a very small time interval, respectively. These are formally defined as

$$m(x) = \lim_{\delta_t \to 0} \frac{E(x_{t+\delta_t} - x_t \mid x_t = x)}{\delta_t} \quad \text{and} \quad v(x) = \lim_{\delta_t \to 0} \frac{E(x_{t+\delta_t}^2 - x_t \mid x_t = x)}{\delta_t}.$$

As usual, diffusion processes for allele frequencies are typically obtained by setting $v(x) = x(1 - x)/(2N_e)$ (the per-generation variance in the change of allele frequencies due to drift), where N_e is the effective size of the population. Then the scale function G is defined by the infinite integral $G(x) = \exp[-2 \int_0^x (m(y)/v(y))\, dy]$.

Kimura [6] showed that the fixation probability of A at the initial frequency p is

$$u(p) = \frac{\int_0^p G(x)\, dx}{\int_0^1 G(x)\, dx}$$

and the expected conditional time to fixation at the initial frequency p is

$$\bar{t}_F = \frac{2(1 - u(p))}{u(p)} \int_0^p \frac{u(x) \int_0^x G(t)\, dt}{v(x)G(x)}\, dx + 2 \int_p^1 \frac{u(x) \int_x^1 G(t)\, dt}{v(x)G(x)}\, dx \tag{1.4}$$

(see the articles by Kimura and Ohta [7] and Walsh and Lynch [8] for other related results).

In this paper, the natural selection assumes one autosomal locus with two alleles, A and a. The three diploid genotypes AA, Aa and aa have different fitnesses, denoted by w_{11}, w_{12} and w_{22}, respectively. These fitnesses are assumed to be constant across generations. Suppose that initially, that is, before selection has operated, the zygote genotypes are in Hardy–Weinberg equilibrium [2], and the frequencies of the A and a alleles are x and y, respectively, where $x + y = 1$. The zygotes then grow to adulthood and reproduce, giving rise to a new generation of offspring zygotes. Our task is to compute the frequencies of A and a in the second generations. As usual, let us denote these by x' and y', respectively, where $x' + y' = 1$.

2. Results

2.1. Relaxing the weak selection assumption In reality, natural selection does not always satisfy the assumption of weak selection. A question that arises is whether or not we can establish analogues of Kimura's [6] results without the assumption of weak selection. The following results present an affirmative answer.

2.1.1 *Deterministic formulae* We first present analogues of Kimura's equations (1.1), (1.2) and (1.3), by removing the extra weak selection assumption as follows.

If $h = 1/2$,

$$t(x, x_t) = -2\left(\ln \frac{1 - x_t}{1 - x} - \frac{1 - s}{s}\right)\left(\ln \frac{x_t}{x} - \ln \frac{1 - x_t}{1 - x}\right) = \frac{2}{s} \ln \frac{x_t(1 - x)}{x(1 - x_t)} - 2 \ln \frac{1 - x_t}{1 - x}. \tag{2.1}$$

If $h = 1$,

$$t(x, x_t) = \frac{1}{s} \ln \frac{x_t}{x} + \frac{1 + s}{s}\left(\ln \frac{1 - x}{1 - x_t} + \frac{1}{x - 1} - \frac{1}{x_t - 1}\right). \tag{2.2}$$

If $h = 0$,

$$t(x, x_t) = \frac{1 + s}{s} \ln \frac{1 - x}{1 - x_t} + \frac{1}{s}\left(\ln \frac{x_t}{x} + \frac{1}{x} - \frac{1}{x_t}\right). \tag{2.3}$$

For the details, see purifying selection in the next subsection and Theorem 2.1 below.

2.1.2 *Stochastic formulae* For the fitnesses $1 + s$, $1 + s/2$ and 1, by using the weak selection assumption, Kimura obtained $G(x) = e^{-2Nsx}$, $u(x) = (1 - e^{-2Nsx})/(1 - e^{-2Ns})$ and

$$\bar{t}_F(p) = \frac{2}{s} \frac{1 - u(p)}{u(p)} \int_0^p \frac{(1 - e^{-2Nsx})^2 e^{2Nsx}}{(1 - e^{-2Ns})x(1 - x)} \, dx$$
$$+ \frac{2}{s} \int_p^1 \frac{(1 - e^{-2Nsx})(1 - e^{-2Ns(1-x)})}{(1 - e^{-2Ns})x(1 - x)} \, dx. \tag{2.4}$$

TABLE 1. Definitions.

	AA	Aa	aa	Total
Fitness	w_{11}	w_{12}	w_{22}	
Frequency after selection	x^2	$2xy$	y^2	1
Frequency after selection	$w_{11}x^2$	$2w_{12}xy$	$w_{22}y^2$	$\overline{W} = w_{11}x^2 + 2w_{12}xy + w_{22}y^2$
Normalized	$w_{11}x^2/\overline{W}$	$2w_{12}xy/\overline{W}$	$w_{22}y^2/\overline{W}$	1

Without the weak selection restriction, Wright [9] showed that $G(x) = (1 + sx)^{-2N}$. Thus,

$$\bar{t}_F(p) = \frac{2(1 - u(p))}{su(p)} \int_0^p \frac{2Nu(x)\{(1 + sx)^{2N} - (1 + sx)\}}{(2N - 1)x(1 - x)}\, dx$$
$$+ 4N \int_p^1 \frac{u(x)\{(1 + sx)^{1-2N} - (1 + s)^{1-2N}\}}{s(2N - 1)x(1 - x)(1 + sx)^{-2N}}\, dx. \tag{2.5}$$

Both equations (2.4) and (2.5) will be used to estimate the error deterministic formulae produced.

2.2. More deterministic formulae   Now we consider Table 1.

Starting with the Hardy–Weinberg equilibrium [2], first we have

$$\Delta x = x' - x$$
$$= \frac{w_{11}x^2 + w_{12}x(1 - x) - x(w_{11}x^2 + 2w_{12}xy + w_{22}y^2)}{(w_{11}x^2 + 2w_{12}xy + w_{22}y^2)}$$
$$= \frac{x(1 - x)[w_{11}x + w_{12}(1 - 2x) - w_{22}(1 - x)]}{(w_{11}x^2 + 2w_{12}xy + w_{22}y^2)}$$
$$= \frac{x(1 - x)[(w_{11} - 2w_{12} + w_{22})x + w_{12} - w_{22}]}{(w_{11} - 2w_{12} + w_{22})x^2 + 2(w_{12} - w_{22})x + w_{22}}$$

and then we have the following theorem.

THEOREM 2.1. *Let the time required for the frequency of A to move from some value x to a value x_t be denoted by $t(x, x_t)$ or t. Then we have the following results.*

(1) *If $w_{11} - 2w_{12} + w_{22} = 0$ (fitnesses are additive), then*

$$t(x, x_t) = -2 \ln \frac{1 - x_t}{1 - x} + \frac{2w_{22}}{w_{12} - w_{22}}\Big(\ln \frac{x_t}{x} - \ln \frac{1 - x_t}{1 - x}\Big).$$

(2) *If $w_{11} - 2w_{12} + w_{22} \neq 0, w_{22} - w_{12} = 0$ (recessive favoured), then $w_{11} - w_{12} \neq 0$ and*

$$t(x, x_t) = \Big(\frac{w_{22}}{w_{11} - w_{12}} + 1\Big) \ln \frac{1 - x}{1 - x_t} + \frac{w_{22}}{w_{11} - w_{12}}\Big(\ln \frac{x_t}{x} + \frac{1}{x} - \frac{1}{x_t}\Big).$$

TABLE 2. Popular selection models [4, Table 3.5].

	AA	Aa	aa
General fitness	w_{11}	w_{12}	w_{22}
(a) Recessive lethal	1	1	0
(b) Detrimental alleles			
(1) Recessive	1	1	$1 - s$
(2) Additive	1	$1 - s/2$	$1 - s$
(c) General dominance			
(1) Purifying selection	1	$1 - hs$	$1 - s$
(2) Positive selection	$1 + s$	$1 + hs$	1
(d) Heterozygote advantage	$1 - c_1$	1	$1 - c_2$
(e) Heterozygote disadvantage	$1 + c_1$	1	$1 + c_2$

(3) *If $w_{11} - 2w_{12} + w_{22} \neq 0$, $w_{11} - w_{12} = 0$ (dominant favoured), then $w_{12} - w_{22} \neq 0$ and*

$$t(x, x_t) = \ln \frac{1 - x}{1 - x_t} - \frac{1}{x_t - 1} + \frac{1}{x - 1}$$
$$+ \frac{w_{22}}{w_{12} - w_{22}} \left(\ln \frac{x_t}{x} - \ln \frac{1 - x_t}{1 - x} + \frac{1}{x - 1} - \frac{1}{x_t - 1} \right). \qquad (2.6)$$

(4) *If $w_{11} - 2w_{12} + w_{22} \neq 0$, $w_{11} - w_{12} \neq 0$ and $w_{22} - w_{12} \neq 0$, then*

$$t(x, x_t) = \frac{w_{22}}{w_{12} - w_{22}} \ln \frac{x_t}{x} - \frac{w_{11}}{w_{11} - w_{12}} \ln \frac{1 - x_t}{1 - x}$$
$$+ \frac{w_{12}^2 - w_{11}w_{22}}{(w_{12} - w_{22})(w_{11} - w_{12})} \ln \frac{w_{12} - w_{22} + (w_{11} - 2w_{12} + w_{22})x_t}{w_{12} - w_{22} + (w_{11} - 2w_{12} + w_{22})x}.$$
$$(2.7)$$

PROOF. The proof of this theorem is routine and is therefore omitted.   □

Note that the time function introduced above is additive. By Theorem 2.1, we are able to give more explicit time-required formulae. Table 2 contains the most popular selection models. We will present explicit time-required formulae for all of the cases.

Case 1. Selection against recessive (Table 2 (b)(1)) By equation (2.6),

$$t(x, x_t) = \ln \frac{1 - x}{1 - x_t} - \frac{1}{x_t - 1} + \frac{1}{x - 1}$$
$$+ \frac{w_{22}}{w_{12} - w_{22}} \left[\ln \frac{x_t}{x} - \ln \frac{1 - x_t}{1 - x} + 1/(x - 1) - 1/(x_t - 1) \right]$$

and, hence,

$$t(x, x_t) = \frac{1 - s}{s} \ln \frac{x_t}{x} - \frac{1}{s} \left(\ln \frac{1 - x_t}{1 - x} - \frac{1}{1 - x_t} + \frac{1}{1 - x} \right).$$

Recall that the result corresponding to Kimura's one for $h = 0$, or with the approximation $1 - sy^2 \approx 1$ of Hartl and Clark [4, (6.15)], is

$$t = \frac{1}{s}\left[\frac{1}{y} - \frac{1}{y_t} + \ln\frac{y(1 - y_t)}{y_t(1 - y)}\right].$$

Case 2. Additive selection (Table 2(b)(2), where the values of fitness are $1, 1 - s/2$ and $1 - s$.) By equation (2.6),

$$t(x, x_t) = -2\ln\frac{1 - x_t}{1 - x} + \frac{w_{22}}{w_{12} - w_{22}}\left(\ln\frac{x_t}{x} - \ln\frac{1 - x_t}{1 - x}\right).$$

If $w_{11} = 1 + s, w_{12} = (2 + s)/s, w_{22} = 1$, then we have equation (2.1).

If $h \neq 1$ and $h \neq 0$,

$$t(x, x_t) = \frac{1 - s}{s - sh}\ln\frac{x_t}{x} - \frac{1}{sh}\ln\frac{1 - x_t}{1 - x} + \frac{1 - 2h + h^2 s}{(s - sh)h}\ln\frac{1 - h + (2h - 1)x_t}{1 - h + (2h - 1)x}. \quad (2.8)$$

As expected, it can be shown that when h approaches zero, the right-hand side of equation (2.8) approaches equation (2.3), and when h approaches 1, the right-hand side of equation (2.8) approaches equation (2.2). This is because the integrand, regarded as a function of h, is continuous on the compact subset $[0, 1]$ and hence is uniformly continuous.

Case 3. Positive selection (Table 2(c)(2), where the values of fitness are $1 + s, 1 + hs$ and 1.) If $h = 1/2$ or $h = 1$, we already have equations (2.1) and (2.2). Now let $h \neq 0$ and $h \neq 1$. By equation (2.7) in Theorem 2.1,

$$t(x, x_t) = \frac{1 + s}{s - sh}\ln\frac{1 - x_t}{1 - x} + \frac{1}{sh}\ln\frac{x_t}{x} + \frac{sh^2 + 2h - 1}{sh - sh^2}\ln\frac{h - (2h - 1)x_t}{h - (2h - 1)x}. \quad (2.9)$$

As in the case of positive selection, it can be shown that when h approaches 1, the right-hand side of equation (2.9) approaches the right-hand side of equation (2.2).

Case 4. Heterozygote advantage (Table 2(d), where the values of fitness are $1 - s_1, 1$ and $1 - s_2$.) By equation (2.7),

$$t(x, x_t) = \frac{1 - s_2}{s_2}\ln\frac{x_t}{x} + \frac{1 - s_1}{s_1}\ln\frac{1 - x_t}{1 - x}$$
$$+ \frac{s_1 s_2 - s_1 - s_2}{s_1 s_2}\ln\frac{s_2 - (s_1 + s_2)x_t}{s_2 - (s_1 + s_2)x}.$$

Case 5. Heterozygote disadvantage (Table 2(e), where the values of fitness are $1 + s_1, 1$ and $1 + s_2$.) By equation (2.7),

$$t(x, x_t) = \frac{1 + s_2}{-s_2}\ln\frac{x_t}{x} - \frac{1 + s_1}{s_1}\ln\frac{1 - x_t}{1 - x}$$
$$+ \frac{s_1 s_2 + s_1 + s_2}{s_1 s_2}\ln\frac{s_2 - (s_1 + s_2)x_t}{s_2 - (s_1 + s_2)x}.$$

2.3.   Estimation of the error deterministic formulae produced   The deterministic formulae cannot predict an accurate result due to the two assumptions, namely, (1) the population size is infinite; (2) assuming that the generation time scale is infinitesimal so that we could use the integral operator. For the second source of errors, we could use a directional computer iteration.

For stochastic estimation, we will focus on the additive fitness case. Recall that the expected conditional time to fixation is calculated at the initial frequency p as

$$\bar{t}_F = \frac{2(1 - u(p))}{u(p)} \int_0^p \frac{u(x) \int_0^x G(t)\,dt}{v(x)G(x)}\,dx + 2 \int_p^1 \frac{u(x) \int_x^1 G(t)\,dt}{v(x)G(x)}\,dx,$$

where

$$G(x) = \exp\left[-2 \int_0^x \frac{m(y)}{v(y)}\,dy\right], \quad u(p) = \frac{\int_0^p G(x)\,dx}{\int_0^1 G(x)\,dx} \quad \text{and} \quad v(x) = \frac{x(1 - x)}{2N}.$$

If fitnesses are $1 + s, 1 + s/2$ and 1, then note that

$$G(x) = [1 + sx]^{-2N} \quad \text{and} \quad u(x) = \frac{1 - (1 + sx)^{-2N+1}}{1 - (1 + s)^{-2N+1}}.$$

Let s be fixed; then equation (2.5) is reduced to

$$\begin{aligned}
\bar{t}_F(p) = {} & \frac{2(1 - u(p))}{su(p)} \int_0^p \frac{2Nu(x)\{(1 + sx)^{2N} - (1 + sx)\}}{(2N - 1)x(1 - x)}\,dx \\
& + 4N \int_p^1 \frac{u(x)\{(1 + sx)^{1-2N} - (1 + s)^{1-2N}\}}{s(2N - 1)x(1 - x)(1 + sx)^{-2N}}\,dx.
\end{aligned} \tag{2.10}$$

Now $\bar{t}_F$ can be regarded as a function of N, and denote $t(p, N) = \bar{t}_F(p)$. Then, for $p_2 \geq p_1$, write $t(p_1, p_2, N) = t(p_1, N) - t(p_2, N)$. Recall the deterministic equation (2.1):

$$t(x, x_t) = \int_x^{x_t} \frac{2(1 + sx)}{sx(1 - x)}\,dx = \frac{2}{s} \ln \frac{x_t(1 - x)}{x(1 - x_t)} - 2 \ln \frac{1 - x_t}{1 - x}.$$

We will prove that

$$\lim_{N \to \infty} t(p_1, p_2, N) = t(p_1, p_2). \tag{2.11}$$

This result gives us some indication that both deterministic and stochastic processes may be correct theories to reveal complicated biological phenomena. We prove it as follows: (1) first we prove that the first term in equation (2.10) vanishes when $\alpha \to \infty$; (2) then we prove that the second term in equation (2.10) approaches

$$\frac{2}{s} \int_x^{x_t} \frac{1 + sx}{x(1 - x)}\,dx = \frac{2}{s} \ln \frac{x_t(1 - x)}{x(1 - x_t)} - 2 \ln \frac{1 - x_t}{1 - x}.$$

Equation (2.11) allows us to estimate the error of deterministic formulae.

Similarly, if the left-hand side of equation (2.11) takes Kimura's function $G(x) = e^{-2Nsx}$, then the right-hand side is that of equation (1.1). That is,

$$\lim_{N \to \infty} t(p_1, p_2, N) = \frac{2}{s}\left[\ln \frac{(1 - p_1)p_2}{(1 - p_2)p_1}\right].$$ (2.12)

Once again this result assures us that both deterministic and stochastic processes are correct theories.

The above procedure can also be applied to a more general selection model, where the fitnesses are $1 + s, 1 + sh$ and 1. Assuming weak selection, we have $G(x) = e^{-2N_e[2shx+s(1-2h)x^2]}$ and, in general, $G(x) = [1 + 2shx + s(1 - 2h)x^2]^{-2N}$. Substituting $G(x)$ into the left-hand side of equation (2.12), we are still able to prove that the right-hand-side function is equation (2.9). That is,

$$\lim_{N \to \infty} t(x, x_t, N) = \frac{1 + s}{s - sh} \ln \frac{1 - x_t}{1 - x} + \frac{1}{sh} \ln \frac{x_t}{x}$$
$$+ \frac{sh^2 + 2h - 1}{sh - sh^2} \ln \frac{h - (2h - 1)x_t}{h - (2h - 1)x}.$$

2.4. Remarks on fixation probability of mutant genes In the literature, much attention has been given to calculating a newly arisen favourable recessive mutation ($h = 0$), at the initial rate $x = 1/2N$ and $N = N_e$ (effective population size). Previously, Haldane [3] used the theory of branching processes to obtain an approximation of $u = \sqrt{s/N} \approx 1.41 \sqrt{s/(2N)}$, while Wright [9] obtained an approximation of $u = 1.1 \sqrt{s/(2N)}$. Kimura [6] obtained $u = \sqrt{2s/(N\pi)}$ for large N. Since $\sqrt{2s/(N\pi)} = 1.31 \sqrt{s/(2N)}$, we see that Kimura's result lies between those of Haldane [3] and Wright [9] and hence is a good approximation: $1.1 \sqrt{s/(2N)} \le \sqrt{2s/(N\pi)} \le \sqrt{s/N}$.

However, all those results work only when selection pressure is not small. The reason is that there is a well-known result for no selection action (only pure drift action): $u = 1/2N$. But when the selection coefficient s approaches zero (meaning no selection action), the above results do not approach $1/2N$. For this reason, we present the following equation:

$$u = \frac{\sqrt{\alpha}}{2N \sqrt{\pi}\{\mathcal{N}_{0,1}(\sqrt{2\alpha}) - 0.5\}},$$ (2.13)

where $\alpha = 2Ns$ and $\mathcal{N}_{0,1}$ is the standard accumulative normal distribution function. Equation (2.13) is not hard to prove by considering the approximation $\sqrt{2\alpha}\mathcal{N}_{0,1}(\sqrt{2\alpha}p) - 0.5 = \int_0^p e^{-\alpha x^2}\, dx$, which is $\approx \sqrt{\alpha}p$ when p is small.

To get Kimura's result, suppose that α is not small; for example, let $\alpha \ge 10$. We have $\mathcal{N}_{0,1}(\sqrt{2\alpha}p) \approx 1$ and, if $p = 1/2N$, then

$$u = u\left(\frac{1}{2N}\right) = \frac{\sqrt{\alpha}}{2N \sqrt{\pi}(1 - 0.5)} = \sqrt{\frac{2s}{N\pi}}.$$

Note that equation (2.13) is valid for all α. When $\alpha \ge 3$, our formula reduces to a simpler form of Kimura's result, $u = \sqrt{2s/(N\pi)}$. For example, when $\alpha = 3$, our result gives $u = 1/N$, while Kimura's formula yields $\sqrt{3/\pi}(1/N)$.

From equation (1.4), there is an exact solution

$$u\left(\frac{1}{2N}\right) = \frac{\int_0^{1/2N} e^{-2Nsx^2}\,dx}{\int_0^1 e^{-2Nsx^2}\,dx}.$$

Now we calculate the fixation probability of a newly arisen allele for the more general case, as described in equation (1.4). Kimura [6, (5.9)] showed that if $0 < h \ll 1$, then

$$u\left(\frac{1}{2N}\right) = \frac{e^{\alpha h^2/(1-2h)}\,\sqrt{2s(1-2h)/N\pi}}{1 - 2\Phi\,\sqrt{2h^2\alpha/(1-2h)}},$$

where $\Phi(x) = (1/\sqrt{2\pi})\int_0^x e^{-t^2/2}\,dt$. We can prove that this equation holds for $0 < h < 1/2$.

2.5. Corrections to Kimura's remark   For the case when $h = 1$, Kimura [6, p. 897] remarked that "For the completely dominant gene ($D = 2h - 1 = 1$) with small selection advantage $s > 0$, we may use the formula $u = 2s$, unless Ns is small". Here we show that this comment may not be correct.

We will consider the case of $h > 1/2$, which is a bit more general. In fact,

$$\int_0^p e^{-2sN_e 2hx + 2sN_e(2h-1)x^2}\,dx = e^{\alpha h^2/(2h-1)}\int_0^p e^{[\alpha h^2/(2h-1)][1-(2h-1)x/h]^2}\,dx$$

$$= e^{\alpha hc}c\int_{1-p/c}^1 e^{\alpha hcy^2}\,dy,$$

where $c = h/(2h - 1)$. Hence, we further have

$$u(p) = \frac{\int_{1-p/c}^1 e^{\alpha hcy^2}\,dy}{\int_{1-1/c}^1 e^{\alpha hcy^2}\,dy} = \frac{\int_0^1 e^{\alpha hcy^2}\,dy - \int_0^{1-p/c} e^{\alpha hcy^2}\,dy}{\int_0^1 e^{\alpha hcy^2}\,dy - \int_0^{1-1/c} e^{\alpha hcy^2}\,dy}$$

$$= \frac{\int_0^{\sqrt{\alpha hc}} e^{y^2}\,dy - \int_0^{(1-p/c)\sqrt{\alpha hc}} e^{y^2}\,dy}{\int_0^{\sqrt{\alpha hc}} e^{y^2}\,dy - \int_0^{(1-1/c)\sqrt{\alpha hc}} e^{y^2}\,dy}$$

$$\to \frac{e^{\alpha hc} - e^{(1-p/c)^2\alpha hc}}{e^{\alpha hc} - e^{(1-1/c)^2\alpha hc}} = \frac{1 - e^{[1-(1-p/c)^2]\alpha hc}}{1 - e^{[1-(1-1/c)^2]\alpha hc}} \to 1$$

when α (or αhc) $\to \infty$. This means that whatever value the initial frequency is, when the population size approaches infinity, the fixation probability approaches 1.

If $h = 1$,

$$u(p) = \frac{\int_{1-p}^1 e^{\alpha y^2}\,dy}{\int_0^1 e^{\alpha y^2}\,dy}.$$

Then, for any initial frequency p, $u(p) \to 1$ when $Ns \to \infty$, which implies that the conclusion above made by Kimura may not be correct.

3. Conclusion

We presented some explicit results for various most general deterministic models of a general type without the weak selection restriction. Theorem 2.1 presents the most general and unified results. Since any deterministic formula always presents an approximation, we also estimated the errors produced by the deterministic formulae, which are important in practice. We have shown that the integral used in deducing these formulae does not produce a significant error. In particular, we have estimated that for the population of over 10^5 and for values of the selection coefficient larger than 0.05, the error produced by the deterministic model is negligible. In addition, we have examined the result of Kimura on fixation probability of a newly arisen mutant. For a newly arisen favourable recessive mutation, we have established a formula which generalizes several results of Kimura [6], Haldane [3] and Wright [9]. For a newly arisen allele for the more general case as described in equation (1.4), we have removed an unnecessary restriction made by Kimura. Also, for a newly arisen favourable dominant mutation, we have indicated why Kimura's comment need not always be correct.

Acknowledgements

References

[1] J. F. Crow and M. Kimura, *An introduction to population genetics theory* (Harper and Row, New York, 1970).

[2] W. J. Ewens, *Mathematical population genetics I: theoretical introduction*, 2nd edn (Springer, New York, 2004).

[3] J. B. S. Haldane, "A mathematical theory of natural and artificial selection. Part V: Selection and mutation", *Proc. Cambridge Philos. Soc.* **23** (1927) 838–844; doi:10.1016/S0092-8240(05)80010-2.

[4] D. Hartl and A. Clark, *Principles of population genetics*, 3rd edn (Sinauer Association, Sunderland, MA, 1997).

[5] P. W. Hedrick, *Genetics of population*, 4th edn (Jones and Bartlett, Burlington, MA, 2011).

[6] M. Kimura, "Some problems of stochastic processes in genetics", *Ann. Math. Stat.* **28** (1957) 882–901; doi:10.1214/aoms/1177706791.

[7] M. Kimura and T. Ohta, "The average number of generations until fixation of a mutant gene in a finite population", *Genetics* **61** (1969) 763–773; http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC1212239&blobtype=pdf.

[8] B. Walsh and M. Lynch, "Evolution and selection of quantitative traits", *Draft version*, 2011; http://nitro.biosci.arizona.edu/zbook/NewVolume_2/newvol2.html.

[9] S. Wright, "Evolution in Mendelian populations", *Genetics* **16** (1931) 97–159; doi:10.1016/S0092-8240(05)80011-4.