A compact result for the time-dependent probability of fixation at a neutral locus

D. Waxman

A compact result for the time-dependent probability of fixation at a neutral locus

D. Waxman

PII: S0022-5193(11)00031-2
DOI: doi:10.1016/j.jtbi.2011.01.019
Reference: YJTBI6331

To appear in: *Journal of Theoretical Biology*

www.elsevier.com/locate/yjtbi

Received date: 26 November 2010
Revised date: 13 January 2011
Accepted date: 13 January 2011

Cite this article as: D. Waxman, A compact result for the time-dependent probability of fixation at a neutral locus, *Journal of Theoretical Biology*, doi:10.1016/j.jtbi.2011.01.019

13/1/2011

A compact result for the time-dependent probability of fixation at a neutral locus

D. Waxman

Faculty of Science and Engineering

University of Brighton

Brighton BN2 4GJ, UK

Running Head: Time-dependent probability of fixation

Correspondence to:

Professor D. Waxman, Faculty of Science and Engineering, Cockroft Building, University of Brighton, Brighton BN2 4GJ, UK.

E-mail: dwaxman.biology@gmail.com

Abstract

A result is derived, in the form of a sum, for the time-dependent probability of fixation of an unlinked neutral locus. The result captures many of the key features of the probability of fixation in a highly compact form. For 'small' times ($t \lesssim 4N_e$) a single term of the sum accurately determines the time-dependent probability of fixation. This is in contrast to the well-known result of Kimura, which requires the contribution of many terms in a different sum, for 'small' times. Going beyond small times, an approximation is derived for the time-dependent probability of fixation which applies for all times when the initial relative allele frequency is small.

Keywords: gene fixation, random genetic drift, diffusion approximation, neutral theory, Wright-Fisher model

1 Introduction

Fixation of an allele in a finite population is a random process. It is characterised by the probability that the allele has fixed by a given time; this is the time-dependent probability of fixation. In the present work we investigate this quantity for an unlinked neutral locus. The standard result for the time-dependent probability of fixation for this case originates with Kimura (1955a), who analysed a Wright-Fisher model (Fisher, 1930; Wright, 1931) under the diffusion approximation.

The diffusion approximation is an approach which was introduced into population genetics by Fisher (1922) and Wright (1945), and then substantially developed by Kimura (1955a). Under a diffusion approximation the relative frequency of an allele is treated as continuous random quantity. The diffusion approximation derives its name from the diffusion equation that governs the distribution of the relative frequency. The diffusion approximation of the Wright-Fisher model leads to an explicit form for the time-dependent probability of fixation that has good accuracy (see e.g., Figure 1 of McKane and Waxman, 2007). However, the diffusion result is a sum of an infinite number of terms (see Eq. (2), below) and this has a complexity that precludes much insight into its behaviour or mathematical form. Furthermore, its determination can require a significant amount of numerical calculation.

Despite the fundamental importance of Kimura's result for the time-dependent

probability of fixation, it is very hard to find many direct applications of it in the literature. In part this may be due to the absence of approximations that summarise its essential aspects in a simple formula for all values of the time. Approximations do, however, exist for times that are relatively long (of the order of the effective population size or larger) and have been employed by Charlesworth *et al.* (2005) in the detection of shared and ancestral polymorphisms; this work thus constitutes a rare application of Kimura's result.

In the present work we reanalyse the formula for the time-dependent probability of fixation and obtain results which capture some of the key features of this quantity in a compact form. These results generally give insights into the dynamics of random genetic drift associated with fixation, and constitute a concrete mathematical handle of the phenomenon which can be approximated to provide explicit formulae. The results also lead to a substantial simplification in the computation of the time-dependent probability of fixation for the case of small and intermediate times.

2 Background

Before we consider details of previous results and the results of the present work, it is useful to introduce a scaled time, τ, which simplifies many expressions. It is defined

in terms of the actual time, t, via

$$\tau = t/(4N_e) \tag{1}$$

where N_e is the effective population size. The quantity τ measures time in units of $4N_e$ generations and henceforth shall simply be referred to as the 'time'.

The general problem under consideration involves a single locus of a randomly mating diploid population with two alleles, denoted A and a. Given that at time $\tau = 0$ the relative frequency of the A allele is p, the probability that the A allele has fixed by time τ (i.e., the time-dependent probability of fixation) is written $P_{\rm fix}(\tau; p)$. For an unlinked locus that is selectively neutral, and in the absence of mutations, the time-dependent probability of fixation was calculated by Kimura under the diffusion approximation, with the result

$$P_{\rm fix}(\tau; p) = p\left[1 - \sum_{n=0}^{\infty} \frac{2(1-p)(2n+3)}{(n+1)(n+2)}(-1)^n C_n^{(3/2)}(1-2p)e^{-(n+1)(n+2)\tau}\right] \tag{2}$$

(Kimura, 1955b). Here $C_n^{(3/2)}(y)$ denotes a Gegenbauer polynomial in the variable y of order 3/2 and degree n (Abramowitz and Stegun, 1965).

The time-dependent probability of *loss* of the A allele, namely $P_{\rm loss}(\tau; p)$, can be obtained from the result of Eq. (2) with the replacement $p \to 1 - p$. This follows

since (i) when the A allele starts at a relative frequency of p, the a allele starts at a relative frequency of $1-p$, (ii) loss of the A allele is equivalent to fixation of the a allele, (iii) under selective neutrality, the two alleles are interchangeable, and hence $P_{\rm loss}(\tau;p) = P_{\rm fix}(\tau;1-p)$. Because of this relation we need consider only the fixation probability.

The expression derived by Kimura for the fixation probability can be approximated by including a finite number of terms in the sum in Eq. (2), but this may require substantial computation. In the presence of selection, techniques have been developed to deal with this (Wang and Ranala, 2004). Generally, the number of terms required in the sum in Eq. (2) depends primarily on the value of the time, τ, since the time-dependent exponentials in Eq. (2), namely $e^{-(n+1)(n+2)\tau}$, only become small for $(n+1)(n+2)\tau \gg 1$. For large n this relation suggests that the number of terms that should be included in the sum is of the order of $\tau^{-1/2}$ and this need not be small if τ is small. For example in a population of $10,000$ individuals, to accurately approximate the fixation probability after 2000 generations can in some cases require at least 20 terms in the sum[1]. Furthermore, the detailed way the expression

[1]We set $p = 0.4$ for an indication (but not a systematic analysis) of the number of terms required in Kimura's sum for $P_{\rm fix}(\tau;p)$, Eq. (2). Including 10 terms in the sum yields a negative value of $P_{\rm fix}(\tau;p)$, while 18 terms yields a 12% error compared with the converged value of the sum (arising from 100 terms). However, including 20 or 21 terms yields $\sim 1\%$ error. We note that 20, numerically, corresponds to $5\tau^{-1/2}$, i.e., is consistent with the estimate that $O(\tau^{-1/2})$ terms are required in the sum.

for $P_{\text{fix}}(\tau; p)$ in Eq. (2) varies with time is not apparent from Kimura's result, except where the time-dependent exponentials in Eq. (2) become small, i.e., at moderately large values of τ. In this case $P_{\text{fix}}(\tau; p) = p\,[1 - 3(1-p)e^{-2\tau} + O(e^{-6\tau})]$.

For values of τ that are moderately small ($\tau \lesssim 1$) Kimura's result can be evaluated numerically. However, no explicit mathematical form, beyond that of the sum in Eq. (2), has so far been given for the probability of fixation for this range of τ, which corresponds to an appreciable set of actual times, ranging from 0 to of the order of $4N_e$ generations.

3 Results

Let us now consider the main result of the present work. This is a formula for the time-dependent probability of fixation, $P_{\text{fix}}(\tau; p)$, that takes a very different form to Kimura's result, Eq. (2). This new formula provides a substantial amount of information about the probability of fixation as a function of time, especially for small τ. The formula presented here is, like Eq. (2), a sum and in Appendix A it is shown that we can write

$$P_{\text{fix}}(\tau; p) = 4\pi^{-1/2} e^{\tau/4} \tau^{-3/2} \sum_{n=0}^{\infty} (-1)^n A_n(\tau; p). \tag{3}$$

Before we give the general form for the $A_n(\tau;p)$ which appear in Eq. (3), we shall present results for a case of practical interest, namely for small initial relative frequencies ($p \ll 1$). For this case, all $A_n(\tau;p)$ are, to leading order in p, directly proportional to p. In particular (see Appendix A),

$$A_n(\tau;p) \simeq p\left(\pi^2/4\right)(2n+1)\, e^{-[(2n+1)\pi]^2/(4\tau)} \tag{4}$$

with corrections of order p^2. Thus for small p we have $A_0(\tau;p) \simeq p\,(\pi^2/4)\, e^{-\pi^2/(4\tau)}$ and keeping just this leading term in the sum in Eq. (3) leads to the explicit approximation

$$P_{\text{fix}}(\tau;p) \simeq p\pi^{3/2} e^{\tau/4} \tau^{-3/2} e^{-\pi^2/(4\tau)}. \tag{5}$$

In Figure 1 we plot the approximation in Eq. (5) for $p = 0.1$ for a range of τ and for comparison also plot the full diffusion result given in Eq. (2).

Figure 1

From Figure 1 we conclude that for small p, Eq. (5) constitutes an explicit approximation of Eq. (2) that: (i) is qualitatively correct over a range of τ, (ii) has small absolute errors, and (iii) when the probability of fixation is appreciable compared with these errors, the approximation is quantitatively correct.

An important feature of the form of $P_{\rm fix}(\tau;p)$ given in Eq. (3) is that the smaller the value of τ, the smaller the number of terms that need to be included in the sum for good accuracy (this is precisely the opposite behaviour to Kimura's result in Eq. (2)). To illustrate this for the case of small p, we note that $A_1(\tau;p)/A_0(\tau;p) \simeq 3e^{-2\pi^2/\tau}$ and for all τ less than 2 (i.e., for all $t < 8N_e$) this ratio is less than 0.02%. Thus, for this range of times, the inclusion of $A_1(\tau;p)$ in the sum in Eq. (3) makes a tiny correction to the leading ($n = 0$) term. It follows that for such 'small' τ, the contribution of just $A_0(\tau;p)$ that was used in Eq. (5) is an approximation that encapsulates an extremely large number of terms of the sum in Eq. (2). More generally, we note that $A_n(\tau;p)/A_0(\tau;p) \simeq (2n+1)e^{-n(n+1)\pi^2/\tau}$ and this ratio rapidly decreases with n, even for moderate τ. For example, for $\tau = 5$ the ratios with $n = 0$, 1, 2 or 3 are approximately $\{1, 6\times10^{-2}, 4\times10^{-5}, 4\times10^{-10}\}$.

Given the level of agreement of the full diffusion approximation, Eq. (2), and the small p approximation of Eq. (5) (see Figure 1) we investigate a further approximation in Figure 2, where the sum representing the diffusion approximation (Eq. (2)) is truncated to the leading two terms: $P_{\rm fix}(\tau;p) \simeq p\,[1-3(1-p)e^{-2\tau}]$ (cf. Charlesworth *et al.*, 2005) however since we are working in a small p approximation, where only the leading p dependence is kept, it is consistent to omit the quadratic p

dependence in this expression. Hence we use

$$P_{\rm fix}(\tau;p) \simeq p\left(1-3e^{-2\tau}\right). \tag{6}$$

Figure 2

From Figure 2 it is apparent that for $2 \leq \tau \leq 3$ there is substantial agreement between the approximation of the present work in Eq. (5) and the result derived from Kimura's analysis, Eq. (6). Additionally, for $\tau \geq 2$ and $p = 0.1$ the approximation of Eq. (6) is very close to the full diffusion result of Eq. (2): the difference is less than 10^{-3}.

We note that the approximation of $P_{\rm fix}(\tau;p)$ in Eq. (5) has the feature that it equals p at the time $\tau = \pi \simeq 3.14$ and, furthermore, beyond this value of τ the approximation exceeds p, which is the largest value that $P_{\rm fix}(\tau;p)$ can take. Including higher terms in the sum of Eq. (3) beyond just the $n = 0$ term pushes this and related features to larger values of τ (see Appendix B), however, a simple way to proceed is to use Eq. (5) for $\tau < 2$ and Eq. (6) for $\tau \geq 2$. At $\tau = 2$ this approximation of Eq. (6) differs from the result of Eq. (5) by less than 0.02%. As a result, a small p

approximation that works for all τ is

$$P_{\rm fix}(\tau;p) \simeq \begin{cases} p\pi^{3/2}e^{\tau/4}\tau^{-3/2}e^{-\pi^2/(4\tau)}, & 0 \le \tau < 2 \\ \\ p\left(1-3e^{-2\tau}\right), & \tau \ge 2. \end{cases} \tag{7}$$

It is worth noting that the dominant factor in Eq. (7), for small τ, is $e^{-\pi^2/(4\tau)}$ which decreases in the vicinity of $\tau = 0$ extremely rapidly. It is primarily this factor alone which leads to $P_{\rm fix}(\tau;p)$ having a very flat curve in the vicinity of $\tau = 0$ (see Figure 1).

A more sophisticated approximation in the 'small' τ range, $\tau < 2$, could be obtained for general p by using the full result for the $A_n(\tau;p)$, and not a small p approximation, and possibly taking more of the $A_n(\tau;p)$ into account. In Appendix A it is shown that the $A_n(\tau;p)$ can generally be written for all p and τ as

$$A_n(\tau;p) = \int_{-\arcsin(\sqrt{p})}^{\arcsin(\sqrt{p})} e^{-(x+n\pi+\pi/2)^2/\tau}\left(x+n\pi+\pi/2\right)\sqrt{p-\sin^2(x)}dx. \tag{8}$$

While this expression appears complex, it is a well-behaved integral that can be straightforwardly evaluated, numerically. Thus, if required, we can obtain essentially exact numerical results for the time-dependent probability of fixation by numerically

evaluating a number of the $A_n(\tau;p)$ and using them in Eq. (3). In Table 1 we illustrate how the truncated approximation, $P_{\rm fix}(\tau;p) \simeq 4\pi^{-1/2}e^{\tau/4}\tau^{-3/2}\sum_{n=0}^{m-1}(-1)^n A_n(\tau;p)$, approaches its true value, when the full form of the $A_n(\tau;p)$ (Eq. (8)) are used, and the number of terms in the sum, m, is increased. We have taken a time of $\tau = 2$ (corresponding to an actual time of $t = 8N_e$) in the calculation of Table 1. Smaller values of τ lead to an even more rapid convergence than that shown in Table 1.

Table 1

It is apparent from Table 1 that the leading term in the sum in Eq. (3), using the numerically calculated value of $A_0(\tau;p)$, is sufficient to determine the full fixation probability to high accuracy for all p, for time $\tau = 2$ (and, indeed, for all smaller times).

In addition to Eq. (3) providing an approximation for $P_{\rm fix}(\tau;p)$, a closely related quantity is the distribution of the random time to fixation, $T_{\rm fix}$, given that fixation ultimately occurs. The probability that $T_{\rm fix}$ is smaller than t is given by $\text{Prob}(T_{\rm fix} < t) = P_{\rm fix}(\tau;p)/p$ thus Eq. (3) can also provide analytical approximations to $\text{Prob}(T_{\rm fix} < t)$ and related quantities and such as the probability density of $T_{\rm fix}$.

4 Summary

In this work, the time-dependent probability of fixation, $P_{\text{fix}}(\tau; p)$, has been derived in the form of a sum. This takes a different form to the result of Kimura, and only a single term of the sum is required, in the regime of 'small' times ($t \lesssim 4N_e$), to approximate the time-dependent probability of fixation. This result has been combined with an approximation of Kimura's result to yield an approximation for $P_{\text{fix}}(\tau; p)$ that holds for small initial relative allele frequencies but all times.

Acknowledgements

I thank Brian Charlesworth for a helpful discussion and two anonymous reviewers for constructive comments that have improved this work.

Appendix A: Proving equivalence with Kimura's result

In this appendix we shall prove that the form for the time-dependent probability of fixation of the present work, Eqs. (3) and (8), is equivalent to Kimura's result, Eq. (2). Proceeding in this direction, i.e., from the result of this work to Kimura's result yields a *much shorter* calculation than proceeding in the opposite direction (which was the way the calculation was originally carried out).

We begin with Eqs. (3) and (8) which can be combined as a sum from $-\infty$ to ∞ as

$$P_{\rm fix}(\tau;p) = \frac{2e^{\tau/4}\tau^{-3/2}}{\sqrt{\pi}} \sum_{n=-\infty}^{\infty} (-1)^n \int_{-\arcsin(\sqrt{p})}^{\arcsin(\sqrt{p})} e^{-[x+(n+1/2)\pi]^2/\tau}$$

$$\times [x+(n+1/2)\pi] \sqrt{p-\sin^2(x)}dx. \tag{9}$$

A key part of the proof involves transforming a sum appearing within Eq. (9), into a different sum, using the Poisson summation method (Apostal, 1979). The result is given in Eq. (11), below. To establish Eq. (11), we note that the sum appearing in

Eq. (9), namely $\sum_{n=-\infty}^{\infty}(-1)^n\left[x+(n+1/2)\pi\right]e^{-[x+(n+1/2)\pi]^2/\tau}$, can be written as

$$\sum_{n=-\infty}^{\infty}(-1)^n\left[x+(n+1/2)\pi\right]e^{-[x+(n+1/2)\pi]^2/\tau}=\int_{-\infty}^{\infty}f(y)\sum_{n=-\infty}^{\infty}(-1)^n\delta(y-n)dy \tag{10}$$

where $f(y)=\left[x+(y+1/2)\pi\right]e^{-[x+(y+1/2)\pi]^2/\tau}$ and $\delta(y)$ is a Dirac delta function. We then note that $\sum_{n=-\infty}^{\infty}(-1)^n\delta(y-n)$ is a function of y that is periodic, with period 2, and accordingly can be written as a Fourier series. We find $\sum_{n=-\infty}^{\infty}(-1)^n\delta(y-n)=\sum_{m=-\infty}^{\infty}e^{i(2m+1)\pi y}$ and using this result in Eq. (10) yields, on evaluating the integral over y,

$$\sum_{n=-\infty}^{\infty}(-1)^n\left[x+(n+1/2)\pi\right]e^{-[x+(n+1/2)\pi]^2/\tau}$$

$$=\frac{\tau^{3/2}e^{-\tau/4}}{2\sqrt{\pi}}\sum_{m=-\infty}^{\infty}(-1)^m(2m+1)e^{-m(m+1)\tau-i(2m+1)x}. \tag{11}$$

This result is exact identity, and using it in Eq. (9) yields

$$P_{\text{fix}}(\tau;p) = \frac{1}{\pi}\sum_{m=-\infty}^{\infty}(-1)^m(2m+1)e^{-m(m+1)\tau}\int_{-\arcsin(\sqrt{p})}^{\arcsin(\sqrt{p})} e^{-i(2m+1)x}\sqrt{p-\sin^2(x)}dx$$

$$= \sum_{m=0}^{\infty}(-1)^m(2m+1)e^{-m(m+1)\tau}I_m \tag{12}$$

where

$$I_m = \frac{4}{\pi}\int_0^{\arcsin(\sqrt{p})}\cos\left[(2m+1)x\right]\sqrt{p-\sin^2(x)}dx. \tag{13}$$

In Eq. (12) we have simplified the integral and used the fact that the part of the sum from $m=-\infty$ to $m=-1$ duplicates the part from $m=0$ to $m=\infty$.

It may be verified that

$$I_0 = p. \tag{14}$$

To determine the I_m for $m>0$ we use Eq. (22.10.11) of Abramowitz and Stegun (1965) for the following representation of a Gegenbauer polynomial in the variable $\cos\theta$ of order n and degree α:

$$C_n^{(\alpha)}(\cos\theta) = \frac{2^{1-\alpha}\Gamma(n+2\alpha)}{n![\Gamma(\alpha)]^2}(\sin\theta)^{1-2\alpha}\int_0^{\theta}\frac{\cos\left[(n+\alpha)\phi\right]}{(\cos\phi-\cos\theta)^{1-\alpha}}d\phi. \tag{15}$$

Here $\Gamma(x)$ denotes Euler's gamma function. Taking $\alpha=3/2$, $p=\sin^2(\theta/2)$ and

$n = m - 1$ in Eq. (15) yields

$$C_{m-1}^{(3/2)}(1-2p) = \frac{2^{-1/2}\Gamma(m+2)}{(m-1)![\Gamma(3/2)]^2}(\sin\theta)^{-2}\int_0^\theta \frac{\cos\left[(m+1/2)\phi\right]}{(\cos\phi-\cos\theta)^{-1/2}}d\phi$$

$$= \frac{4m(m+1)}{2^{1/2}\pi}\frac{1}{1-(1-2p)^2}\int_0^{2\arcsin\sqrt{p}} \frac{\cos\left[(m+1/2)\phi\right]}{(\cos\phi-1+2p)^{-1/2}}d\phi \quad (16)$$

and on setting $\phi = 2x$ gives

$$C_{m-1}^{(3/2)}(1-2p) = \frac{2m(m+1)}{\pi p(1-p)}\int_0^{\arcsin\sqrt{p}} \cos\left[(2m+1)x\right]\sqrt{p-\sin^2 x}dx. \quad (17)$$

Comparing this result with Eq. (13) leads to

$$I_m = \frac{2p(1-p)C_{m-1}^{(3/2)}(1-2p)}{m(m+1)}, \qquad m = 1, 2, 3, ... \quad (18)$$

Using Eqs. (14) and (18) in Eq. (12) yields Kimura's result, Eq. (2), hence we have shown that Eqs. (3) and (8) are fully equivalent to Eq. (2).

To derive the leading term in a small p approximation of $A_n(\tau; p)$ we change

variable in Eq. (8) from x to $y = x/\sqrt{p}$. This yields

$$A_n(\tau;p) = p\int_{-\arcsin(\sqrt{p})/\sqrt{p}}^{\arcsin(\sqrt{p})/\sqrt{p}} e^{-\left[\sqrt{p}y+(n+1/2)\pi\right]^2/\tau}\left[\sqrt{p}y+(n+1/2)\pi\right]\sqrt{1-p^{-1}\sin^2(\sqrt{p}y)}dy. \tag{19}$$

Expanding all quantities within the integral leads to $A_n(\tau;p) = p\int_{-1}^{1} e^{-[(n+1/2)\pi]^2/\tau}(n+1/2)\pi\sqrt{1-y^2}dy + O(p^2)$. The remaining integral has the value $\pi/2$ hence for small p we obtain the approximate form of $A_n(\tau;p)$ given in Eq. (4).

Appendix B: Dependence of $P_{\rm fix}(\tau;p)$ on τ when the sum is truncated.

In this appendix we investigate the form of $P_{\rm fix}(\tau;p)$ given in Eq. (3), when it is approximated by truncating the sum at a finite number of m terms, in which case

$$P_{\rm fix}(\tau;p) \simeq 4\pi^{-1/2}e^{\tau/4}\tau^{-3/2}\sum_{n=0}^{m-1}(-1)^n A_n(\tau;p). \tag{20}$$

When m is odd, this is numerically found to lead to an approximation for $P_{\rm fix}(\tau;p)$ which achieves the value of p (which is the maximum value that $P_{\rm fix}(\tau;p)$ can take) at a finite value of τ. Beyond this value of τ, the approximation overshoots p. By contrast, taking m even leads to an approximation for $P_{\rm fix}(\tau;p)$ which achieves a maximum value (less than p) at a finite value of τ, with the approximation decreasing beyond this value of τ. It follows that keeping m terms in the sum, the value of the scaled time τ where one of these two behaviours occurs ($P_{\rm fix}(\tau;p)$ either equalling p or achieving its maximum value) corresponds to the largest value of τ where the approximation can be sensibly applied. Writing this largest value as $\tau_{\rm max}$, we have determined its values for different numbers of terms in the sum, m, using the small p approximation given in Eq. (4). The results are summarised in Table 2.

Table 2

The value of τ adopted for Table 1 (in the main text) is $\tau = 2$ and is smaller than any of the $\tau_{\max}$'s appearing in Table 2. The rapid convergence of the truncated sum of Eq. (20), as demonstrated in Table 1, provides evidence that for values of τ less than $\tau_{\max}$ the truncated sum provides a very good approximation to the fixation probability given in Eq. (2).

There is a straightforward measure of the error in the approximation of Eq. (20) precisely at the 'maximum time' $\tau_{\max}$, namely the difference between the full diffusion result for $P_{\rm fix}(\tau; p)$, when evaluated at $\tau = \tau_{\max}$ and the value of the approximation of Eq. (20), when also evaluated at $\tau = \tau_{\max}$. This measure of the error depends on the initial frequency, p, and we have adopted the value $p = 0.1$ and employed the small p approximation for the $A_n(\tau; p)$. We then find that when $m = 1$, 2 or 3 the absolute values of the differences are approximately $\{5 \times 10^{-4}, 8 \times 10^{-6}, 8 \times 10^{-8}\}$ and hence very small compared with $P_{\rm fix}(\tau_{\max}; p)$ which is very close to p.

References

Abramowitz, M. and Stegun, I., 1965 *Handbook of Mathematical Functions.* Dover, New York.

Apostal, T. M., 1979 *Mathematical Analysis*, 2nd Edition. Addison-Wesley, Reading, MA.

Charlesworth, B., Bartolome, C. and Veronique, N. 2005 The detection of shared and ancestral polymorphisms. Genet. Res., Camb. 86: 149–157.

Fisher, R. A., 1922 On the dominance ratio. Proc. R. Soc. Edinburgh 42: 321–431

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford.

Kimura, M., 1955a Stochastic processes and distribution of gene frequencies under natural selection. Cold Spring Harbour Symp. Quant. Biol. 20: 33–53.

Kimura, M., 1955b. Solution of a process of random genetic drift with a continuous model. Proc. Natl. Acad. Sci. 41: 141–150.

McKane, A. J. and Waxman, D., 2007 Singular solutions of the diffusion equation of population genetics. J. Theor. Biol. 247: 849-858.

Wang, Y. and Ranala, B., 2004 A novel solution for the time-dependent probability of gene fixation or loss under natural selection. Genetics 168: 1081–1084.

Wright, S. 1931 Evolution in Mendelian populations. Genetics 16: 97–159.

Wright, S. 1945 The differential equation of the distribution of gene frequencies. Proc. Natl Acad. Sci. USA 31: 382–389.

Figure Captions

Figure 1 Caption. With p the initial relative frequency and τ the scaled time, $\tau = t/(4N_e)$, this figure illustrates a 'small p' approximation to the time-dependent probability of fixation of Eq. (5). The approximation arises from including just the small p form of the $n = 0$ term in Eq. (3). The approximation is plotted against time, τ (solid line). In the same figure the full diffusion result (obtained by including 100 terms in the sum in Eq. (2)) is also plotted (broken line). The value of p adopted for the figure is $p = 0.1$ and for the range of τ adopted for the figure ($0 \leq \tau \leq 3$) the difference between the results of Eqs. (2) and (5) is less than 6×10^{-3}. Smaller values of p lead to yet closer agreement. For example, for $p = 0.01$ and the same range of τ, the difference is less than 6×10^{-5}.

Furthermore, for $p = 0.1$ and $0 \leq \tau \leq 3$ (the values adopted for the figure), it is not possible to visually distinguish between the full diffusion result, Eq. (2), and the approximation derived from Eq. (3) where only the *full* zeroth order term (not its small p approximation) is included in the sum; the results differ by less than 5×10^{-4}.

Figure 2 Caption. In this figure the small p approximation to $P_{\text{fix}}(\tau; p)$, given by Eq. (5) is compared with the result of keeping just the leading two terms of the diffusion sum, when quadratic dependence on p is omitted, namely Eq. (6). The initial frequency used in the figure was $p = 0.1$.

Over the range $1 \leq \tau \leq 2$ the approximations of Eqs. (5) and (6) differ by less than 1.4×10^{-3}, while over the range $2 \leq \tau \leq 3$ the two approximations differ by less than 5×10^{-4}.

Tables and Table Captions

Table 1

	$p = 0.2$	$p = 0.4$	$p = 0.6$	$p = 0.8$
$m = 1$	0.191259	0.386966	0.587181	0.792055
$m = 2$	0.191211	0.386814	0.586811	0.791206
$m = 3$	0.191211	0.386814	0.586811	0.791206
Diffusion result	0.191211	0.386814	0.586811	0.791206

Table 1. Caption: To illustrate how many terms are required in Eq. (3) to accurately determine the time-dependent probability of fixation, we have calculated the value of $P_{\text{fix}}(\tau; p)$, from Eqs. (3) and (8), when $\tau = t/(4N_e) = 2$ and the sum has been truncated to m terms (the leading term in the sum is A_0 and the highest term is A_{m-1}). When $\tau = 2$, the quantity $4\pi^{-1/2}e^{\tau/4}\tau^{-3/2}$, which appears in Eq. (3), has the value $c = 1.315...$ and the first three entries of the column headed $p = 0.2$ are the values of $P_{\text{fix}}(\tau; p)$ when it is successively approximated by $c \times A_0$, $c \times (A_0 - A_1)$

and $c \times (A_0 - A_1 + A_2)$, with the A_n's numerically evaluated from Eq. (8) at $\tau = 2$ and $p = 0.2$. The final entry in the column headed $p = 0.2$ is the full diffusion result that is calculated from Eq. (2). The other columns in the table follow from similar considerations.

Table 2

No. of terms in the sum, m	1	2	3	4	5	6	7	8
$\tau_{\max}$	3.1	5.3	7.4	9.5	11.7	13.7	15.9	18.0

Table 2. Caption: Truncating the sum in Eq. (3) at m terms leads to the approximation of the time-dependent fixation probability of Eq. (20). This approximation works up to a limited value of the time τ, which we denote $\tau_{\max}$, and which we have estimated from the small p results of Eq. (4). There appears to be a simple linear relationship between $\tau_{\max}$ and m, namely $\tau_{\max} \simeq 2.12m + 1.02$ that we have verified to very reasonably hold for m ranging from 1 to 20.