**Appendix: Derivation and implementation details of the mathematics used in Kimura\_NumIntegr.py**

The mathematics used in program *Kimura\_NumIntegr.py* (numerical integration of Kimura’s diffusion fixation probability[[1]](#endnote-1),[[2]](#endnote-2)) is derived here. To avoid getting lost in too much detail, some readers may wish to skip some detailed derivations, which are shown in grey background.

**1. Wright–Fisher model and diffusion limit**

We consider a diploid Wright–Fisher population of effective size with two alleles and , and genotype fitnesses

(1)

Let denote the frequency of allele in the population. In the Wright–Fisher model, the expected allele frequency in the next generation is

(2)

where is the expected deterministic change due to selection. For weak selection (), and ignoring higher-order terms in , one obtains

(3)

**Derivation of eqn. (3)**

Eqn. (3) is an approximation based on a Taylor expansion in *s* and dropped terms, as will be elaborated on below.

The computer program Kimura\_NumIntegr.py assumes that changes per generation are small, as Kimura did (typical assumption: *Ne* → ∞ while selections *s* → 0 so that *Ne* can be finite).

In practice, diffusion approximations remain useful well beyond infinitely small , *provided that* is large so allele frequency changes are still small from generation to generation. Consequently, ***predicted probabilities of fixing are not precisely reliable by the program when s is large or Ne is small***.

It would be possible instead of the approximation (3) to use in Kimura\_NumIntegr.py the exact discrete-generation expression for *M*(*p*) (i.e., with no Taylor expansion):

(4)

**Details of how eqn. (3) can be derived.**

Start from the standard genotype-fitness model

(5)

genotype frequencies (Hardy–Weinberg) *AA*: *p*2, *Aa*: 2*p*(1−*p*), *aa* = (1−*p*)2 and the allele frequency before reproduction .

The allele frequency after viability selection (expected allele frequency among adults) is

(6)

where the mean fitness is

(7)

and p´ has the same meaning as pt+1.

Compute numerator and denominator explicitly.

Numerator:

Denominator (mean fitness):

Therefore,

(8)

Define by , so

(9)

with

Now expand to first order in (i.e. drop terms). Use  
. (10)

Then

*p*´ ≈ (*p* + *sA*)(1 – *sC*)

*p´* = *p* + *sA* – *spC* + *O*(*s*2),

where *p´* means the value of *p* in the next generation.

Therefore,

(11)

Compute :

(12)

Factor *p*(1–*p*) out:

Therefore, to a first order in *s* this leads to eqn. (3),

The variance in allele frequency change caused by random genetic drift is

(13)

where

**Derivation of eqn. (13)**

In a diploid Wright–Fisher population of effective size , the number of *A* alleles in the next generation, *K*, is drawn from a binomial distribution:

*K* ~ Binomial(2*Ne*,*p*),

where *p* is the current allele frequency.

The variance of a binomial distribution is well known:

Var (*K*) = 2*Nep*(1–*p*).

Since the next generation’s allele frequency is *p´* = *K*/(2*Ne*), we obtain (13), which is the variance in allele frequency change due to random genetic drift:

V(*p*) = Var(*p´*) | *pt* = *p*)

**2. Diffusion approximation and boundary-value problem**

Taking the diffusion limit of the discrete Wright–Fisher process *Ne* → ꚙ, *s* → 0 with *Nes* of finite value and using the per-generation drift *M*(*p*) and variance *V*(*p*) obtained in (3)–(13), the allele-frequency process converges to a continuous diffusion on the interval .  
The time development of its probability density *u*(*p*,*t*) can be described by the Fokker–Planck (forward Kolmogorov) equation shown below as (14).

(14)

For fixation probabilities one can apply the backward Kolmogorov equation,[[3]](#endnote-3) which gives the probability *π*(*p*) will eventually fix, assuming that the allele had an initial frequency 0:

(15)

where

Equation (15) is a second-order ordinary differential equation with absorbing boundaries at π = 0 or 1. This means that once the diffusion process reaches a boundary at 0 or 1 it stays there permanently. This corresponds to an allele *A* of *a* fixing in a population.

**3. Solution via integrating factor**

Equation (15) is a second-order linear homogeneous differential equation of *π*(*p*). To solve it, it can be reduced to a first-order equation in *π´*(*p*) following the standard method for diffusion equations.[[4]](#endnote-4)

We introduce, an integrating factor which is key to finding the general solution for the fixation probability.

(16)

we obtain a total derivative and after integration the general solution for the fixation probability:

(17)

This expression is the fundamental diffusion-theory formula for fixation probability (Kimura, 1957).[[5]](#endnote-5)

**4. Application to genotype model selection**

Substitute M*p* and V*p* from eqn. (3) and eqn. (13) into the exponent of *ψ*(*p*):

(18)

Define the scaled selection parameter

(19)

Then the inner exponent becomes

(20)

Defining , the fixation probability simplifies[[6]](#endnote-6) to

(21)

To make notation more compact, express as

(22)

**5. Closed-form reductions**

**5.1 Semidominant (linear) case**

When , and the exponent becomes linear, with . Then

(23)

and the fixation probability reduces to the semidominant closed form[[7]](#endnote-7)

(24)

Numerically this is evaluated stably using the expm1 function to avoid cancellation when is small:

(25)

The function, expm1(x) = *ex* − 1 is available in numerical computing libraries for many programming languages to solve accurately for very small values of x.

**5.2 Quadratic exponent ().[[8]](#endnote-8)**

For the fixation probability is

(26)

with Completing the square gives

(27)

So

(28)

With the substitution (so ) the integral (28) evaluates to

(29)

Erf, the error function for a complex number *z* is defined as

An analogous expression holds for the denominator with . Because the multiplicative prefactor

is identical in numerator and denominator it cancels, giving the closed form

(30)

When the integral involves is numerically unstable, and numerical integration should be used.

**6. Numerical integration procedure**

**6.1 Exponent shift for numerical stability**

To prevent overflow or underflow in computing , subtract the minimum of Φ(*y*) on [0,1]:

Φmin ​ =min{Φ(0), Φ(1), Φ(*y*∗)},

where

Define the shifted exponent

(31)

and use in the integrals. This preserves numerical accuracy because the constant factor cancels in the ratio defining .

**6.2 Numerical quadrature**

A uniform grid of points with spacing is used. The cumulative integral is approximated by the trapezoidal rule:

(32)

The numerator of Equation (21) is obtained by interpolating *C*(*p0​)* at the desired initial frequency *p0*; the denominator is *C*(1). Thus,

(33)

**7. Numerical accuracy, stability, and convergence**

For *n* subintervals and using the usual nomenclature for the Trapezoid integration algorithm (*h* = (*b−a*)/*n*, where *h* = step size = Δy, *a* = lower bound of integration, *b* = upper bound of integration), the error is *O*(*h*2) = *O*(*n−*2).

For twice continuously differentiable the error bound may be written as:[[9]](#endnote-9)

(34)

Empirical tests indicate that increasing the grid from 5001 to 10001 points changes by less than for typical parameter sets.

1. Kimura, M., Solution of a process of random genetic drift with a continuous model, *Proc. Natl. Acad. Sci. USA*, **41**:144–150, 1955. Available online: <https://www.pnas.org/doi/abs/10.1073/pnas.41.3.144> [↑](#endnote-ref-1)
2. Kimura, M., On the Probability of Fixation of Mutant Genes in a Population, *Genetics* 47:713–719, 1962. Available online: <https://pmc.ncbi.nlm.nih.gov/articles/PMC1210364/> [↑](#endnote-ref-2)
3. Ewens, W. J., Mathematical Population Genetics. I: Theoretical Introduction, 2nd ed. Springer, 2004. See Chapter 4 (“Diffusion theory”) for a careful derivation of the diffusion (Fokker–Planck) equation from the discrete Wright–Fisher model. Available online: <https://archive.org/details/springer_10.1007-978-0-387-21822-9/page/n153/mode/2up> [↑](#endnote-ref-3)
4. Ref 3, section 4.3, in particularly pp. 86–88. Available online: <https://archive.org/details/springer_10.1007-978-0-387-21822-9/page/n155/mode/2up> [↑](#endnote-ref-4)
5. Kimura, M., Some problems of stochastic processes in genetics, *Annals of Mathematical Statistics* **28**, 882–901, 1957. Available online: <https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-28/issue-4/Some-Problems-of-Stochastic-Processes-in-Genetics/10.1214/aoms/1177706791.full> [↑](#endnote-ref-5)
6. According to AI systems Gemini and ChatGPT, this is equivalent to eqn. (4) and (13) in ref 2. [↑](#endnote-ref-6)
7. According to AI systems Gemini and ChatGPT, this is equivalent to eqn. (5) and Equation (19) in ref. 2. [↑](#endnote-ref-7)
8. The equations in this section were solved and confirmed independently with AI tools Claude, Grok, ChatGPT and Gemini. [↑](#endnote-ref-8)
9. The trapezoid method is a well-known and highly accurate method used in numerical analysis to calculate integrals. See for example Atkinson, K.E., *An Introduction to Numerical Analysis*, 2nd Edition, John Wiley & Sons, 1989. [↑](#endnote-ref-9)