



# Spread of an Advantageous Allele through Space and Time



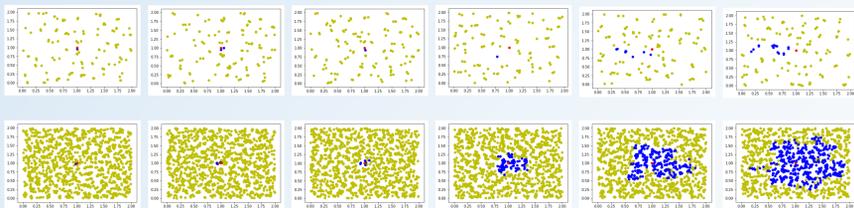
Kyle Shaw and Peter Beerli, Department of Scientific Computing, Florida State University, Tallahassee FL USA

New alleles (genes) occur frequently. Occasionally these new alleles are advantageous to the individual who carries them. When this advantageous allele increases the likelihood of the individual having more offspring, then, according to the basic theory of natural selection and evolution, we would expect the new allele to spread through the population and replace the old allele. The time it takes for this new allele to fixate in the population is an important question for epidemiologists, ecologists, and especially wildlife preserve policy makers. While the topic has been studied for several years, none of the current methods and models for predicting the speed at which a new allele will spread take into account the spatial dimension. In 1937, Ronald Fisher attempted to describe the wave of a new allele travelling through a population with diffusive partial differential equation. However, at the time, his method was computationally intractable and other less accurate, yet computationally feasible methods were explored. I first show, through a simple simulation that ignoring the spatial dimension can lead to invalid conclusions. I then develop a mathematical model that estimates the expected speed of the wave of advance a new advantageous allele would cause.

## Introduction

Assume that  $N$  individuals are uniformly distributed across a geographic space bounded by  $[0, B]^2$ . At each time step each individual has some offspring according to a Poisson Distribution. Each of those offspring are distributed about their parent according to a two dimensional Normal Distribution. Let a new mutation be inserted into an individual near the center of the geographic region. Let this new mutation be advantageous, thus increasing the likelihood of producing offspring. It is expected that when we track the individuals who contain this new mutation that they will increase through time and spread through the geographic space almost like a ripple in a pond.

The selectional advantage ( $s$ ) of a new allele obviously has an affect on the speed ( $v$ ) at which a new allele will spread. It has also been shown that in models without a geographic component that larger population sizes ( $N$ ) increase the effectiveness of selectional advantage. Within a model with a two dimensional component,  $N$  translates to population density ( $\rho$ ). I show below, through simulation, that an increased  $\rho$  leads directly to an increased  $v$ .



These graphs show a little slide show. Each snap shot is after several generations have passed. From the left to the right we see that as we travel forward in time the individuals how contain the advantageous allele (colored blue) spread out from the center. The top row is a sparsely populated area, while the second row is very densely populated. Notice how the spread is much more symmetric in the dense population. This is because randomness has less of an effect. The mutation also tends to spread faster and farther in dense populations.

## Single Random Walk

To predict the  $v$  of a new allele I first model an individual random walk. Each generation an individual is placed around its parent according to a normal distribution  $N(0, \sigma^2)$ . Let  $x_t, y_t$  be the coordinates and  $d_t = \sqrt{x_t^2 + y_t^2}$  be the distance from the origin at generation  $t$ . Thus

$$x_0 = 0 \quad y_0 = 0 \quad (1)$$

$$x_1 = N(0, \sigma^2) \quad y_1 = N(0, \sigma^2) \quad (2)$$

$$x_2 = N(x_1, \sigma^2) \quad y_2 = N(y_1, \sigma^2) \quad (3)$$

$$= x_1 + N(0, \sigma^2) \quad = y_1 + N(0, \sigma^2) \quad (4)$$

$$= N(0, 2\sigma^2) \quad = N(0, 2\sigma^2) \quad (5)$$

$$x_t = N(0, t\sigma^2) \quad y_t = N(0, t\sigma^2) \quad (6)$$

The Probability Density Function  $f()$ , Cumulative Density Function  $F()$  and Expectation  $E()$  of  $d_t$  can be calculated as

$$f(d_t) = \int_0^{2\pi} \frac{1}{2\pi t\sigma^2} e^{-\frac{d_t^2}{2t\sigma^2}} d\theta \quad (7)$$

$$= \frac{1}{t\sigma^2} e^{-\frac{d_t^2}{2t\sigma^2}} \quad (8)$$

$$F(d_t) = \int_0^{d_t} f(r)r \, dr \quad (9)$$

$$= 1 - e^{-\frac{d_t^2}{2t\sigma^2}} \quad (10)$$

$$E(d_t) = \int_0^\infty f(d_t)d_t^2 \, dd_t \quad (11)$$

$$= -d_t e^{-\frac{d_t^2}{2t\sigma^2}} \Big|_0^\infty - \int_0^\infty e^{-\frac{d_t^2}{2t\sigma^2}} dd_t \quad (12)$$

$$= \sigma\sqrt{\frac{t\pi}{2}} \quad (13)$$

Thus the velocity of a single random walk can be calculated as

$$v = \frac{\sigma\sqrt{\frac{t\pi}{2}}}{t} \quad (14)$$

$$= \sigma\sqrt{\frac{\pi}{2t}} \quad (15)$$

## Maximum of k random Walks

The front end of the wave was next modeled as the maximum of  $k$  random walks. Let  $d_i$  be the distance from the origin of the  $i$ th random walk at time  $t$  where  $i = 1 \dots k$ . Let  $r = \max(d_0, \dots, d_k)$ . From the Cumulative Density Function  $F_M()$  of  $r$  we can calculate the PDF  $f_M()$  of  $r$  as follows

$$F_M(x) = P(r < x) \quad (16)$$

$$= P(d_0 < x \wedge \dots \wedge d_k < x) \quad (17)$$

$$= P(d_0 < x) \dots P(d_k < x) \quad (18)$$

$$= P(d_0 < x)^k \quad (19)$$

$$= F(x)^k \quad (20)$$

$$f_M(x) = F'_M(x) \quad (21)$$

$$= kF(x)^{k-1}f(x) \quad (22)$$

$$= \frac{ke^{-\frac{x^2}{2t\sigma^2}}}{t\sigma^2} (1 - e^{-\frac{x^2}{2t\sigma^2}})^{k-1} \quad (23)$$

Which leads to the  $E_M$  of the maximum distance the the expected velocity as

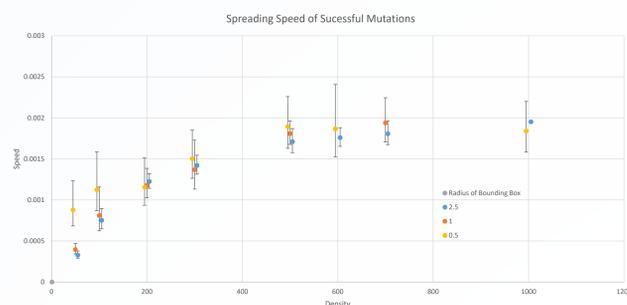
$$E_M = \int_0^\infty f_M(x)x^2 dx \quad (24)$$

$$= \int_0^\infty \frac{kx^2 e^{-\frac{x^2}{2t\sigma^2}}}{t\sigma^2} (1 - e^{-\frac{x^2}{2t\sigma^2}})^{k-1} dx \quad (25)$$

$$= \sigma k \sqrt{\frac{t\pi}{2}} \sum_i^{k-1} \binom{k-1}{i} \frac{-1^i}{(i+1)^{1.5}} \quad (26)$$

$$v = \frac{\sigma k \sqrt{\frac{t\pi}{2}} \sum_i^{k-1} \binom{k-1}{i} \frac{-1^i}{(i+1)^{1.5}}}{t} \quad (27)$$

$$= \sigma k \sqrt{\frac{\pi}{2t}} \sum_i^{k-1} \binom{k-1}{i} \frac{-1^i}{(i+1)^{1.5}} \quad (28)$$



This graph shows the velocity vs Density. Each simulation was run over several geographic ranges as shown in the key. Here we can see two things. First, as density increases, velocity increases as well, with diminishing returns. Second, beyond a certain density, velocity is independent of time.

## Conclusions

We can see from the previous graph that velocity is independent of time. From equation 28 we see a  $\sqrt{t}$  in the denominator. We know that  $k$  changes over time thus the summation in the equation must scale with  $\sqrt{t}$ .

otherwise velocity would not be independent of time. Also worth noting is that increasing density only increases the velocity up so a horizontal tangent. This tangent is determined by the dispersal rate.

## Summary

Population density, not size, affects the rate at which a new allele will spread. The rate at which the number of walks increases is dependent upon both population density and selection advantage. The velocity of the traveling wave is independent of time, thus the function  $F_n$  must scale with  $\sqrt{t}$ .