Journal of Abstract and Computational Mathematics

http://www.ntmsci.com/jacm

Analysis of F-statistics under Selection with Continuous and Discrete Models

Anuraag Bukkuri

University of Minnesota, United States

Received: 8.8.2019, Accepted: 6.12.2019
Published online: 20.12.2019

Abstract: The F-statistic is a statistic which measures shared genetic drift among sets of populations and can be used to test various admixture hypotheses. However, thus far, this statistic has only been developed in the context of genetic drift, and ignores other evolutionary forces such as mutation and selection. This paper examines and further develops the F statistic under models of selection. Specifically, the F-statistic is developed under stochastic PDE models or discrete models with additive selection, random fluctuation of selection intensities, and selection with mutation. These results can help expand the scope of these statistics in theoretical population genetics.

Keywords: F-statistics, Theoretical Population Genetics.

1 Introduction

The F statistic, which measures shared genetic drift among sets of populations has been used extensively in the literature to determine if populations are related in a tree-like manner [Reich et al. 2009], if a population descended from several ancestral populations [Reich et al. 2009], what the closest relative to a population is [Raghavan et al. 2014], how mixtures/splits of a population explain demography [Patterson et al. 2012; Lipson et al. 2013], number of founder populations [Reich et al. 2012; Lazaridis et al. 2014], and contributions from different population to a population of interest [Green et al. 2010; Haak et al. 2015]. A thorough description of the F statistic, different interpretations of it, and relevant derivations for some useful expressions are given in [Peter 2016]. The one drawback of the statistic is that it assumes genetic drift and ignores any other evolutionary scenarios. In this paper, continuous and discrete models capturing changes in gene frequency are created and simulated, before using these models to gain further insight into the F statistic under various evolutionary scenarios.

2 PDE and Discrete Models of Gene Frequency

Here, we create and analyze models of changes in gene frequency under random genetic drift, genic selection without dominance, genic selection with random fluctuation of selection intensities, and genic selection with mutation, in that order.

* Corresponding author e-mail: bukku001@umn.edu

2.1 Random Genetic Drift

A stochastic PDE model with only random genetic drift can be created as follows:

$$\frac{\partial \phi}{\partial t} = \frac{1}{4N}\frac{\partial^2}{\partial x^2}(x(1-x)\phi) \tag{1}$$

in which $\phi = \phi(p,x;t)$ is the probability density that the gene frequency becomes x in the t^{th} generation, given that it is p at $t = 0$. To solve the above, we let $\phi = TX$ and, after substitution into the above equation, we have:

$$\frac{1}{T}\frac{\partial T}{\partial t} = \frac{1}{4N}\frac{\partial}{X\partial x^2}(x(1-x)X) \tag{2}$$

Since both sides of the equation must equal a constant (call this λ), the above equation can be separated into two ODEs:

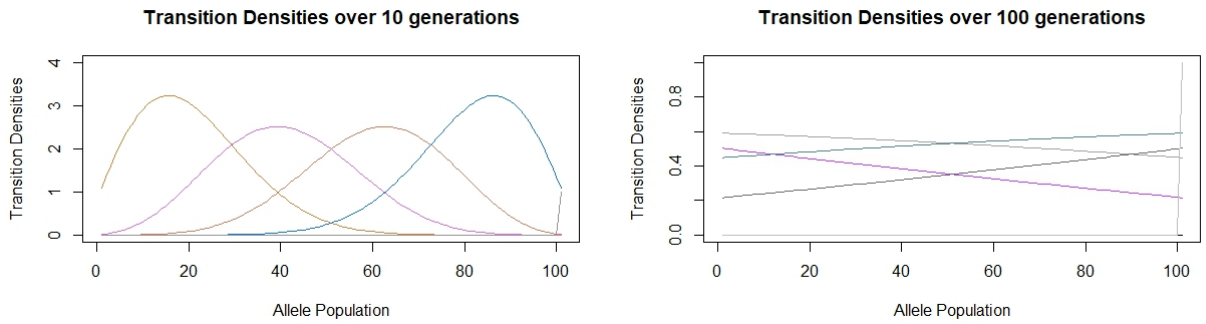
$$\frac{dT}{dt} = -\lambda T \tag{3}$$

$$x(1-x)\frac{d^2X}{dx^2} + 2(1-2x)\frac{dX}{dx} - (2-4N\lambda)X = 0. \tag{4}$$

From this, the solution to the overall PDE reveals itself nicely when one uses the Gegenbauer polynomials (falls out of the latter, hypergeometric equation). Specifically, the solution can be represented as follows:

$$\sum_{i=1}^{\infty} \frac{(2i+1)(1-r^2)}{i(i+1)} T_{i-1}^1(r)T_{i-1}^1(z)exp(-i(i+1)t/4N) \tag{5}$$

where $r = 1-2p$, $z = 1-2x$, and $T_{i-1}^1(n)$ represent the Gegenbauer polynomials. From this, I created simulations depicting the transition densities of an allele with initial frequencies of 0, 0.25, 0.5, 0.75, and 1 after 10 and 100 generations in a population size of 100. For numerical approximation, the infinite sum in equation 6 was instead evaluated until 70.

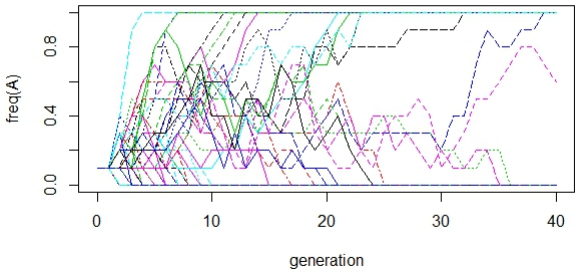


As one can see, there seems to be several binomial distributions in the case with 10 generations, other than when the simulation was run for an initial allele frequency of 0. However, as would be expected, when one increases the generation time to 100, a completely different result is observed. In this case, there are no longer any binomial distributions. The initial allele frequency at 0 corresponds to the bottom, horizontal line corresponding to the transition density at 0. The closer the initial allele frequency is to 0.5, the smaller the magnitude of the respective slope is. Moreover, if the allele frequency is less than 0.5, the slope will be negative, while if the frequency is more than 0.5, its slope will be positive. In

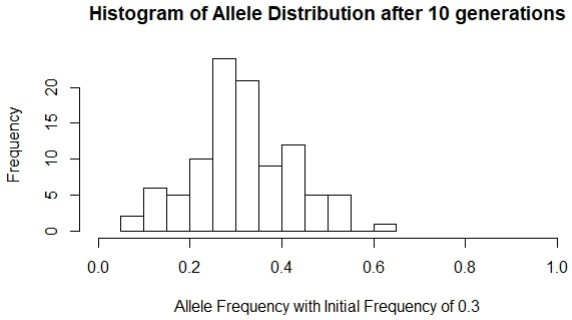
the discrete case of random drift, the equation for the expectation of allele a in a population over time t can be written as the following:

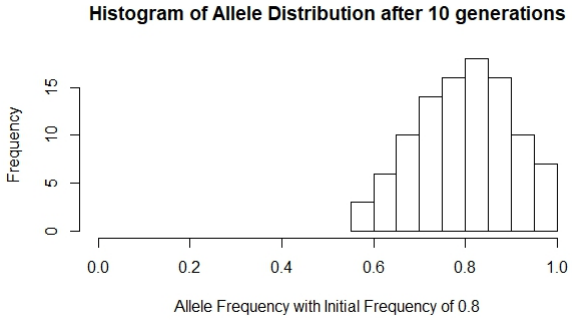
$$\frac{k}{N} \tag{6}$$

since the expected value of the proportion of allele a, under just random genetic drift will not be expected to change as a function of number of generations. A simulation was run using an initial allele a number of 1 in a population of 10. This was run for 40 generations with 100 trials. The simulation results are shown below. Clearly, one can see that the allele frequencies vary randomly, most of them eventually reaching either fixation or loss.



Moreover, distribution plots analogous to the above two transition density graphs can be made for this discrete case. Identical parameters to those above were used. The first plot used 30 initial alleles in a population of 100, and was run for 10 generations with 100 trials. The second plot below used 80 initial alleles out of 100, run for 10 generations with 100 trials. Note that this captures the short-term evolutionary dynamics, so that most alleles do not reach fixation/loss as they did in the above simulation, run for 40 generations.





As one can see, there exists a strong binomial distribution in these discrete cases as well, in accordance with our earlier results.

2.2 Genic Selection without Dominance

A PDE model with superimposed selection can be written as the following:

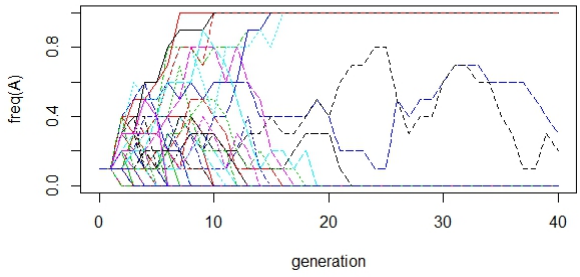
$$\frac{\partial \phi}{\partial t} = \frac{1}{4N}\frac{\partial^2}{\partial x^2}x(1-x)\phi - s\frac{\partial}{\partial x}x(1-x)\phi \tag{7}$$

where $\phi = \phi(p,x,t)$ is the probability density that the gene frequency becomes x in the t^{th} generation given that it's p at t_0, N is the population size, and s represents the selection coefficient.

Note that for a panmictic population of size N, with individuals of types a and b, if the generation at time t has k individuals of type a then, according to the above model, the generation at time t+1 can be described as:

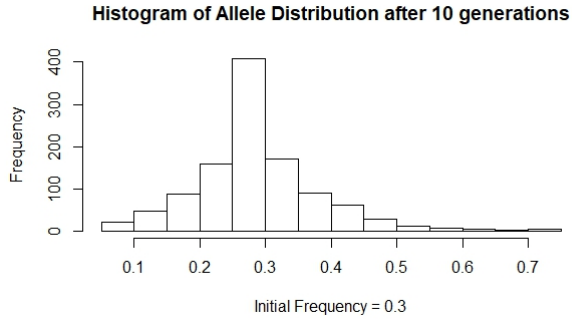
$$\frac{k(1+s)}{k(1+s)+N-k} \tag{8}$$

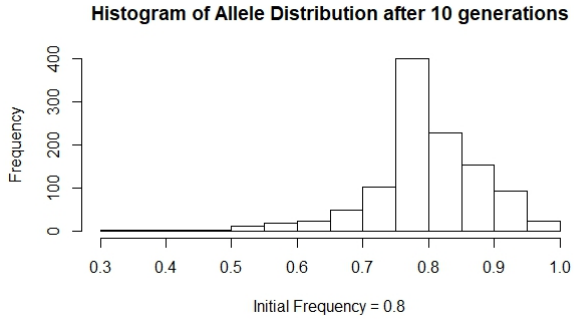
Simulations of this model over time were performed. Specifically, a selection coefficient of 10^{-4}, 1 initial mutant allele in a population of 10 were run over 40 generations. This trial was repeated 100 times; the results can be seen below.



As one can see, x=0 and x=1 are the absorbing states in this model (i.e. the allele is either lost or reaches fixation). The probability that the population under consideration contains both alleles decreases over time as one allele (the one with selective advantage) drifts towards fixation, while the other one is lost. Specifically, this probability decreases at the rate of the smallest eigenvalue of the PDE in equation 2.

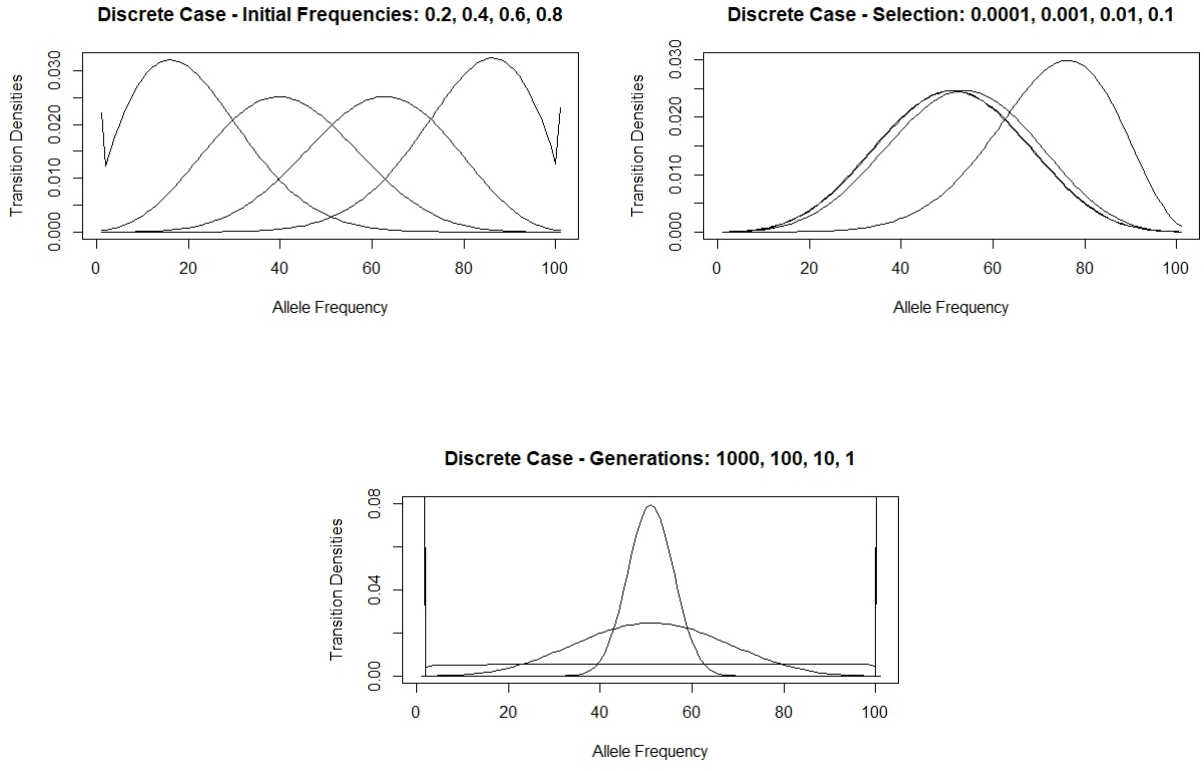
As before, one can also create simulations for transition densities. A selection coefficient of 10^{-4} was used in a total population of 100. 100 trials were run over 10 generations. The first histogram below used an initial mutant population of 30, while the second used an initial population of 80. Again, note that, since we are only considering short-term dynamics, most of the alleles do not reach fixation or loss. In other words, and this will be the case in most of the graphs discussed in this paper, the distribution will be primarily unimodal for a certain number of generations, but will soon switch to a bimodal distribution after that.





When one considers an initial frequency of 0.8, there is still a roughly binomial distribution in the graphs. However, the binomial distributions in both cases is slightly higher than the initial allele frequency. Note that this is the expected result, since selection for this allele was introduced into the population.

Calculations were also made based on this discrete model by using equation 9 to construct a 101x101 stochastic matrix capturing the probabilities from moving from one state to another given the same initial conditions as above: a total population of 100, a selection coefficient of 10^{-4}, run for 10 generations. Three plots were then created from this matrix: one which plotted transition densities for varying initial frequencies with a selection coefficient of 10^{-4}, another which plotted densities for varying selection coefficients with an initial mutant frequency of 0.5, and the last one plotted densities for varying generations with an initial mutant frequency of 0.5 and a selection coefficient of 10^{-4}. The results can be seen below:



As one can see, the same trends that were observed in the histogram can be seen here. In the first graph, it's clear that, for each initial allele frequency, after 10 generations, there existed a binomial distribution around the initial frequency. However, as can be seen in the second graph, when one increases the selection coefficient (drastically), there is still a binomial distribution, but the mode has now shifted to the left, displaying selection at work. In the third graph, one may notice that the greater the number of generations, the more spread out the binomial distribution is. This occurs until the 1000th generation, at which one may see that the mutant frequency reaches either the 0 or 1 absorbing states.

To work with the continuous model, we must first obtain a closed form solution of the PDE. To do this, we first rewrite the above equation as the oblate spheroidal equation:

$$(1-k^2)\frac{d^2V}{dk^2} - 4k\frac{dV}{dk} + (4N\lambda - 2 - c^2) + c^2k^2V \tag{9}$$

for k ∈ (-1,1) = 1-2x. Then, the final analytical solution falls out nicely when one utilizes the Gegenbauer polynomials, a class of orthogonal polynomials that generalize Legendre and Chebyshev polynomials:

$$\phi(p,x,t) = \sum_{n=1}^{\infty} L_n e^{-\lambda_n t + 2lx} C_n(k) \tag{10}$$

where

$$L_n = \frac{(1-z^2)e^{-c(1-z)C_n(z)}}{\sum'_{n=0,1} \frac{(k+1)(k+2)}{2n+3}(f_k^n)^2} \tag{11}$$

for $z = 1-2p_0$, $c = Ns$, and the primed summation means to sum over even values of k if n is even, and over values of k if n is odd and

$$C_n(k) = \sum_{k=0,1}^{'} f_k^n C_k^{\alpha}(k) \tag{12}$$

where f_k^n are constants that can directly be found from [Stratton et al. 1941], the primed summation defined as before. The $C_k^{\alpha}(k)$ represents the Gegenbauer polynomials defined as follows:

$$C_{i-1}^{\alpha}(k) = \frac{i(i+1)}{2} F(i+2, 1-i, 1, \frac{1-k}{2}) \tag{13}$$

where F is the hypergeometric series defined as:

$$F(a,b,2,x) = \frac{\Gamma(2)\Gamma(2-a-b)}{\Gamma(2-a)\Gamma(2-b)} F(a,b,a+b-1,1-x) + \frac{\Gamma(2)\Gamma(a+b-2)}{\Gamma(a)\Gamma(b)} (1-x)^{2-a-b} F(2-a,2-b,3-a-b,1-x) \tag{14}$$

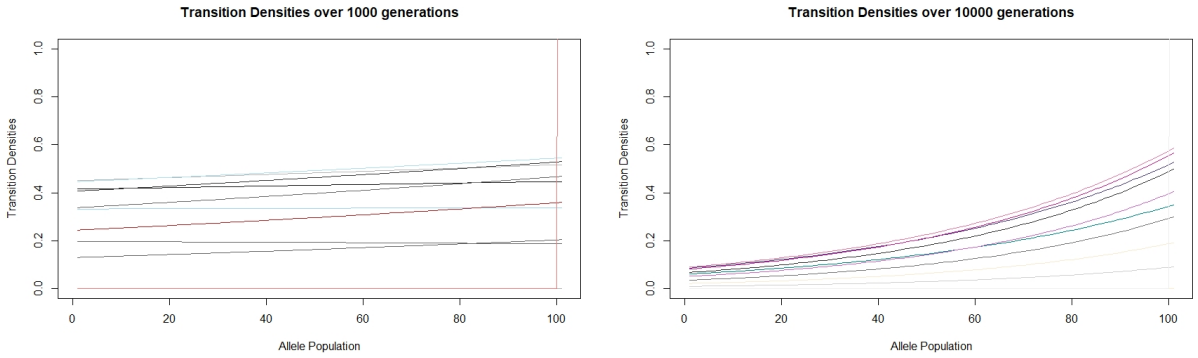
For numerical estimates, only the first three eigenvalues (and thus eigenfunctions and coefficients) of equation 3 were considered in the computations. The first 3 eigenvalues were derived from [Stratton et al. 1941] using the following relation:

$$\lambda_k = \frac{c^2 - B_{1,k}}{4N} \tag{15}$$

where $B_{1,k}$ are the separation constants.

Two calculations were made: the left plot uses a selection coefficient of 10^{-4} and an N of 10^3; the right plot uses a selection coefficient of 10^{-4} and an N of 10^4. The below table shows the values of $B_{1,k}$ and f_k^n used in the calculations:

	c=1	c=0.1
$B_{1,0}$	-1.79531	-1.998
$B_{1,1}$	-5.56753	-5.599571
$B_{1,2}$	-11.53482	-11.99533
f_0^0	1.0208	1.0002
f_2^0	0.01398	0.0001334
f_1^1	1.0314	1.0003
$-f_0^2$	0.03486	0.00034288
f_2^2	0.9898	0.99983



2.3 Genic Selection with Random Fluctuation of Selection Intensities

The random fluctuation of selection intensities can be interpreted as changes in the environment of alleles, which randomly influence selection coefficients of alleles. However, random sampling of gametes can also cause such fluctuations; to remedy this, one can assume an infinite population size. Now, assuming s is the selective advantage of allele A_1 over allele A_2 s.t. the rate of change in the frequency of A_1 for some value of s is sx(1-x). Additionally, we assume that A_1 is, on average selectively neutral. Then, the mean of s over an infinite amount of time will be 0 and its variance, V_s will be constant (depending on how much the selection fluctuates). Then, the following PDE represents our situation:

$$\frac{\partial \phi}{\partial t} = \frac{V_s}{2}\frac{\partial^2}{\partial x^2}(x^2(1-x)^2\phi) \tag{16}$$

for $0 < x < 1$. To solve this, we first convert it to the heat conduction equation by letting the following be true:

$$u = \frac{1}{2}e^{V_s t/8}x^{3/2}(1-x)^{3/2}\phi \tag{17}$$

$$\zeta = log(\frac{x}{1-x}) \tag{18}$$

Then, we have the following PDE:

$$\frac{\partial u}{\partial t} = \frac{V_s}{2}\frac{\partial^2 u}{\partial \zeta^2} \tag{19}$$

for $-\infty < \zeta < \infty$. Thus, now we can solve equation 16 as the following, when the initial distribution of gene frequency is $\phi(x,0)$:

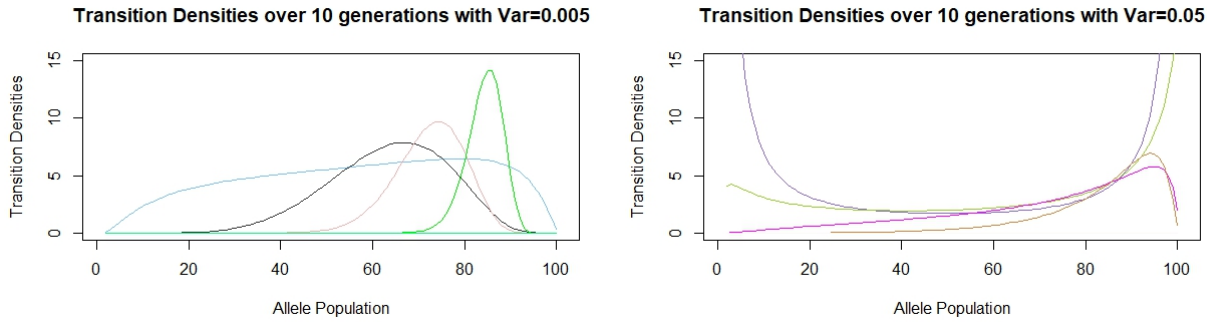
$$\phi(x,t) = \frac{1}{\sqrt{2\pi V_s t}}\frac{e^{V_s t/8}}{x(1-x)^{3/2}}\int_0^1 exp(-\frac{log\frac{x(1-y)}{(1-x)y}^2}{2V_s t})\sqrt{y(1-y)}\phi(y,0)dy \tag{20}$$

However, if our initial condition is just a fixed gene frequency (as will be the case in our simulations), then the above formula can be simplified to the following:

$$\phi(p,x;t) = \frac{1}{\sqrt{2\pi V_s t}}exp(-\frac{V_s}{8}t - \frac{log\frac{x(1-p)}{(1-x)}^2}{2V_s t})\frac{\sqrt{p(1-p)}}{(x(1-x))^{3/2}} \tag{21}$$

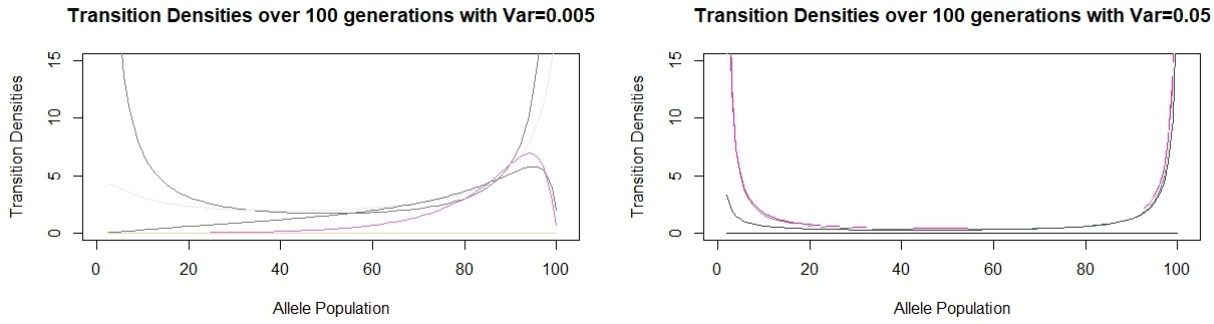
Simulations were performed to examine the gene frequency under random fluctuation of selection intensities. Specifically, initial frequencies of 0, 0.25, 0.5, 0.75, and 1 were used, with a total population of 100. Two selection variances were used as well: one at 0.05 and another at 0.005. The simulations were run for 10 and 100 generations. The results are provided below:

Generations: 10 From this, one can see that for the lower variance, most of the curves are relatively binomial. However,



when one increases the variance (and thus the fluctuation of selection intensities), some of the curves now become bimodal, with the modes at the absorbing states.

Generations: 100 We note that when the generations are increased to 100 *and* the variance is increased, the vast majority



of curves become bimodal.

It's interesting to note that, as can be seen in the PDE equations, time and variance of fluctuation in selection intensities are directly proportional. In our simulations, one can see that the transition densities for 10 generations with a variance of 0.05 and for 100 generations with a variance of 0.005 are identical.

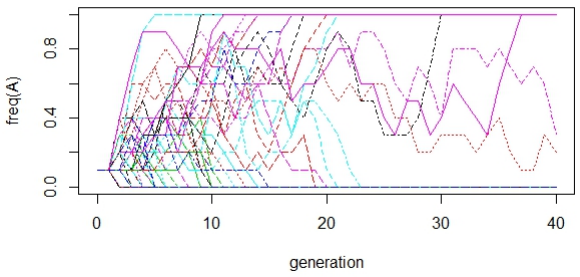
2.4 Genic Selection with Mutation

Similarly, one can derive equations for the allele frequencies in a population under the stochastic PDE model with mutation and selection.

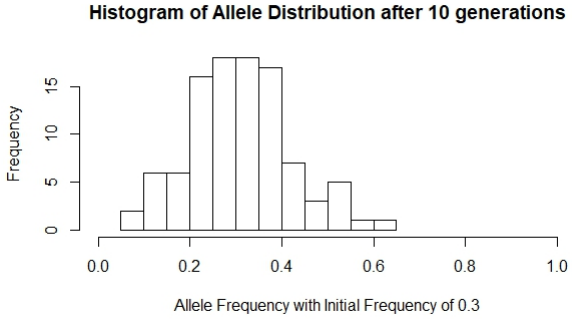
Specifically, assuming a total population of N, k individuals of type a, the rest of type b, ϕ_1 a individuals mutate to b and ϕ_2 b individuals mutate to a, the proportion of their potential offspring that are of type a after selection and mutation can be described as:

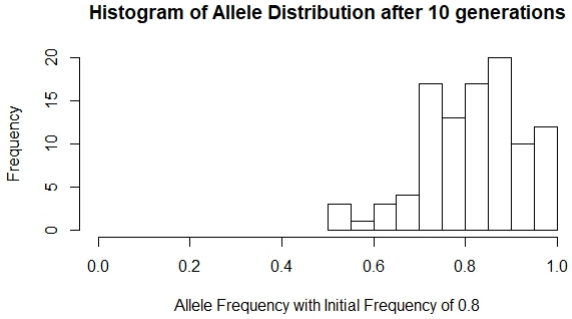
$$\frac{k(1+s)(1-\phi_1)}{k(1+s)+N-k}+\frac{(N-k)\phi_2}{k(1+s)+N-k} \tag{22}$$

As before, simulations are performed to see the results of mutation and selection in practice. The below simulation was run with a selection coefficient (for a) of 10^{-4}, u1 and u2 of 10^{-8}, initial number of a as 1 in a population of 10, over 40 generations with 100 trials. The results can be seen below:



Similar to the PDE model with selection, the vast majority of alleles either reached selection or fixation. Again, we can examine short-term dynamics by simulating the population over the course of 10 generations. Below are two simulations for initial starting populations of 30/100 and 80/100.



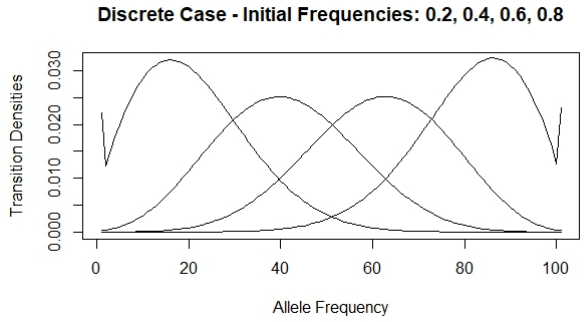


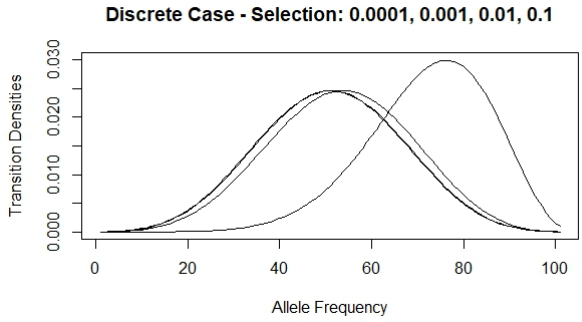
In this case, we can see that though the binomial distribution is still fairly present, there are many other peaks/high frequencies in the graph. This randomness makes sense due to the addition of mutations. Again, we see that the binomial distribution often occurs slightly over the initial allele frequency, thanks to the introduction of selection.

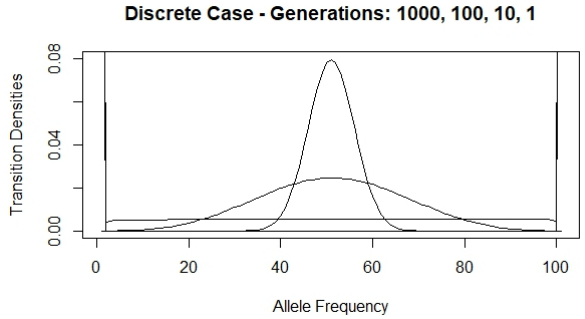
Calculations were again made based on this discrete model by using the same method described in the discrete selection case. 5 plots were then created from this matrix. Unless otherwise described, standard conditions were $u1 = u2 = 10^{-8}$, $s = 10^{-4}$, 10 generations, and 100 initial population size.

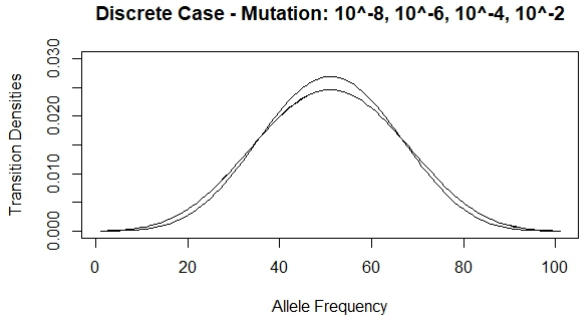
1.Plotted transition densities for varying initial frequencies
2.Plotted densities for varying selection coefficients
3.Plotted densities for varying number of generations
4.Plotted densities for varying mutation rates
5.Plotted densities by letting u2=0 and changing the u1 mutation rate

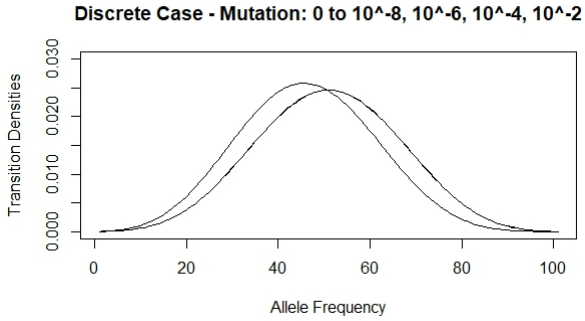
The results can be seen below:











As one can see, the first three graphs here are identical to those in just the selection case, thus following the same trends. When one considers changing the mutation rate, we can see that the higher the mutation rate, the sharper the binomial curve is (i.e. the transition densities are even higher close to the initial allele frequency). Moreover, when we just changed the u1 mutation rate, we can see that (for high enough mutation rates), the allele frequency after 10 generations shifted to the left, favoring the second allele. This makes sense because we allowed for "conversion" from allele 1 to 2, but not the other way around.

3 Genetic Drift: F_2

The F_2 statistic was developed to quantify the genetic drift between two populations. For two populations, P_1 and P_2 and corresponding allele frequencies p_1 and p_2, the F_2 statistic is defined as follows [Peter 2016]

$$F_2(P_1,P_2) = F_2(p_1,p_2) = \mathbb{E}(p_1 - p_2)^2 \tag{23}$$

The F_2 statistic can help describe variances in allelic frequencies, heterozygosity, probability of identity by descent, correlation among individuals, and the probability of coalescence [Peter 2016].

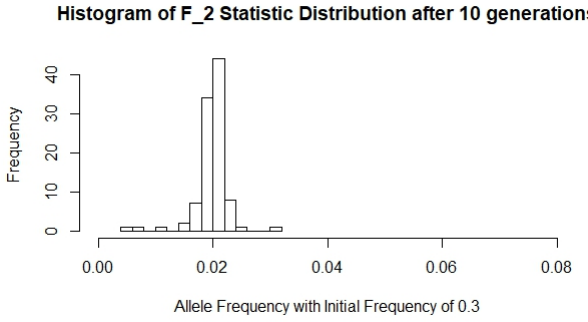
We analyze this statistic in alleles within a single population sampled at times t_0 and t. The corresponding populations are set as P_0 and P_t, respectively. Then, the F_2 statistic under genic selection with no dominance, as well as with both selection and mutation can be calculated as follows:

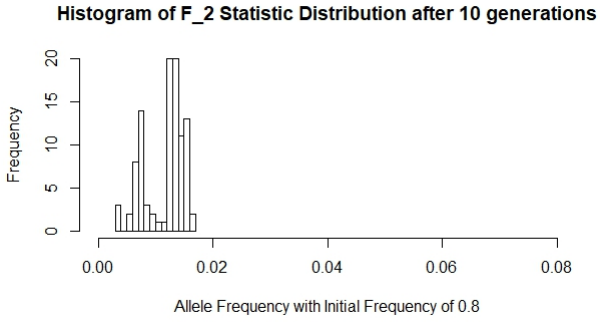
$$F_2(P_t,P_0) = \mathbb{E}(p_t - p_0)^2 \tag{24}$$
$$= Var(p_t - p_0) + \mathbb{E}(p_t - p_0)^2 \tag{25}$$
$$= Var(p_t) + Var(p_0) - 2COV(p_0,p_t) + \mathbb{E}(p_t - p_0)^2 \tag{26}$$

where p_t is as described on the right hand side of equation 4.

Simulations were performed comparing the F_2 statistic in a single population with and without selection, as well as with both selection and mutation. The results can be seen below:

First, the PDE model with just random genetic drift under the discrete case was implemented with identical parameter values as used before (initial numbers of 30 and 80 in an initial population of 100, over 10 generations, with 100 replicates). Histograms capturing the results can be seen below:

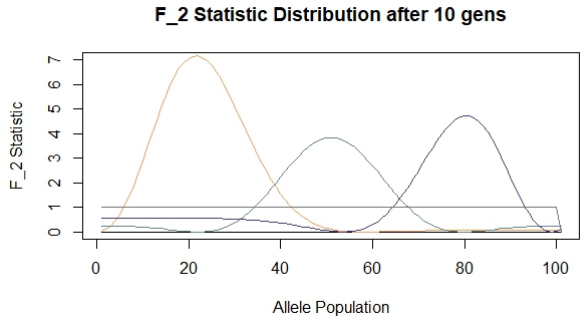


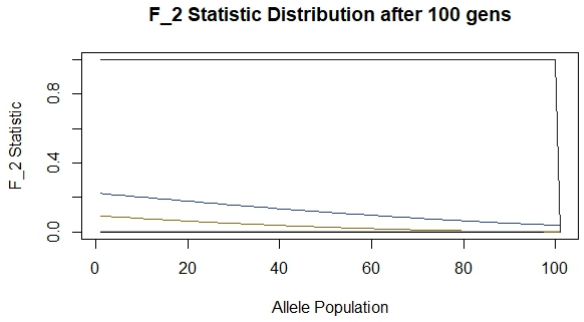


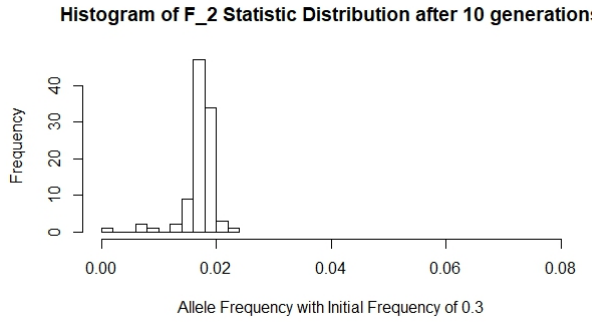
As one can see, the modal F_2 statistic for the initial allele frequency of 0.3 was just past 0.02, whereas the modal F_2 statistic for the initial allele frequency of 0.8 was around 0.015.

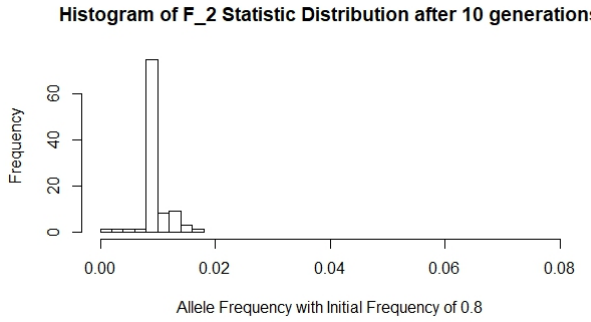
Similarly, distributions can be captured using the continuous PDE models. Specifically, simulations were run to capture the F_2 statistic distribution after 10 and 100 generations using initial frequencies of 0, 0.25, 0.5, 0.75, and 1.
Here, one can see that, in the short-term, the F_2 statistic distribution seems to be evenly split between binomial distribution and simply a straight line. However, when one increases the number of generations to 100, one can see that F_2 statistic distribution has a negative slope other than our end cases.

Next, the discrete version of the model with selection was run. Specifically, a selection coefficient of 10^{-4} was used, with a mutant allele population of 1 in a total population of 10. Time t_0 was set to generation 0, and time t was set to 40 generations; 100 trials were run. The results can be seen below:
In this case, for both initial allele frequencies, the average F_2 statistic decreased. However, the general trend is kept, with the modal F_2 of the 0.3 initial frequency, at 0.018, being higher than the modal F_2 of the 0.8 initial frequency, which was



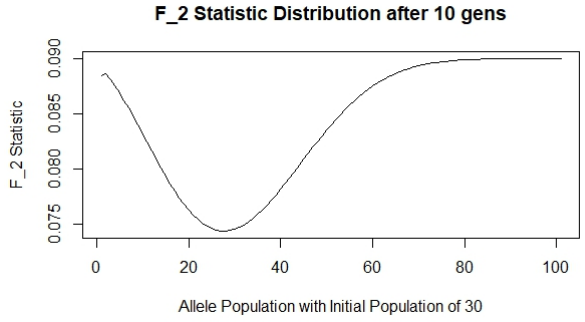


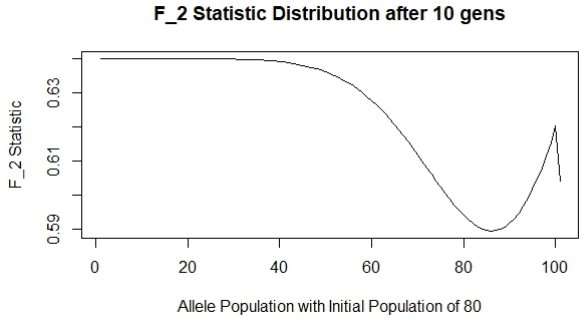




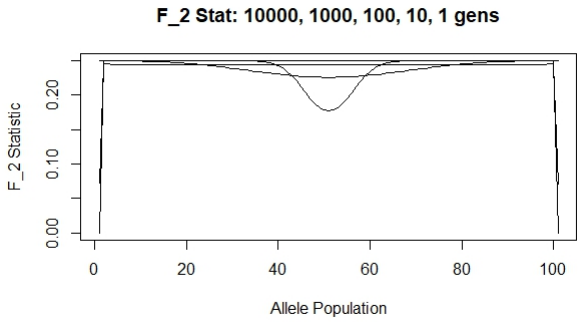
closer to 0.015.

Graphs can also be made using the calculations made earlier. With the same parameters, the following two graphs were produced:

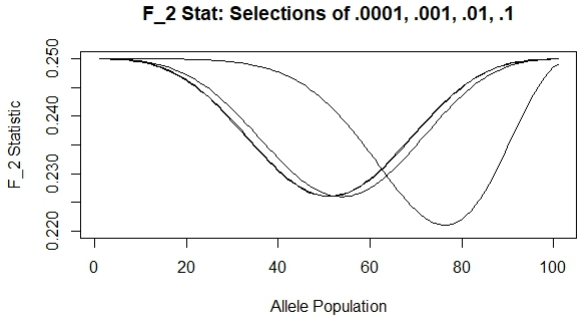




We can see that the F_2 statistic is lowest in both cases for the allele populations closest to the initial mutant population. Another simulation was performed looking at the F_2 statistic under varying numbers of generations. The resulting graph can be seen below:

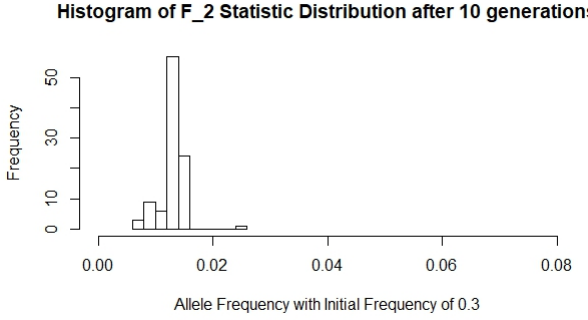


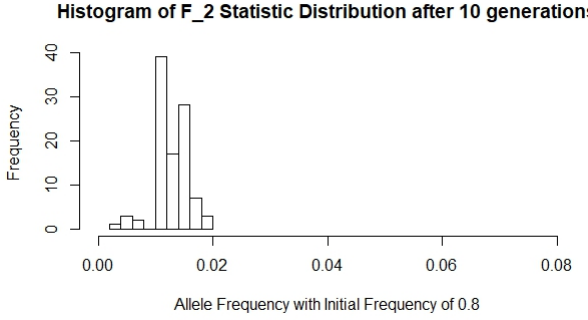
As one can see, the higher the number of generations the trial was run for, the more stable the F_2 statistic was over the population. When the trial was run for only 1 generation, it's clear that the F_2 statistic was much lower near the initial allele frequency of 0.5. Another calculation was made for the F_2 distribution under varying selection:



Here, the F_2 statistic of the curve with the lowest selection coefficient was lowest around 0.5, the initial mutant allele frequency. However, with the highest selection of 0.1, the F_2 statistics curve was noticed to be lowest near 0.8, as a result of the strong selective pressures.

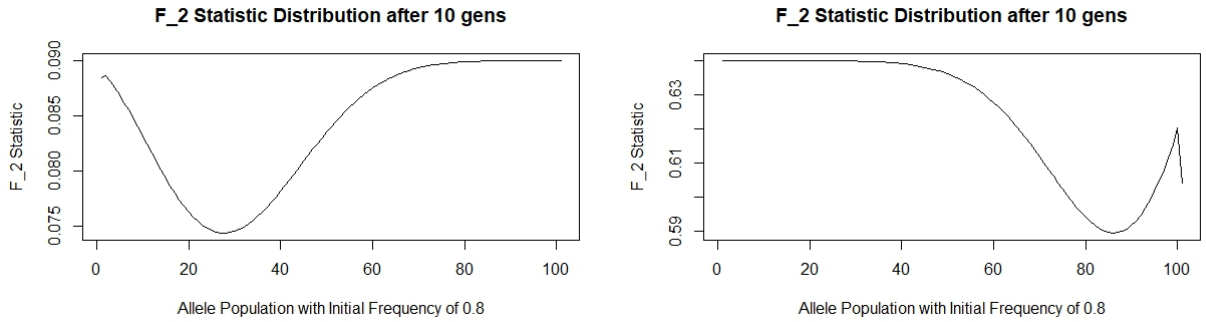
Then, simulations for the discrete PDE model with selection and mutation were performed. A selection coefficient of 10^{-4} was used, with a mutant allele population of 1 in a total population of 10, and both ϕ_1 and ϕ_2 were set to 10^{-8}. This was run for 10 generations, with 100 trials. Two simulation results are provided below:

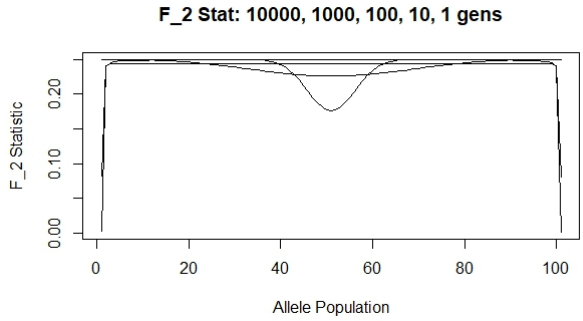


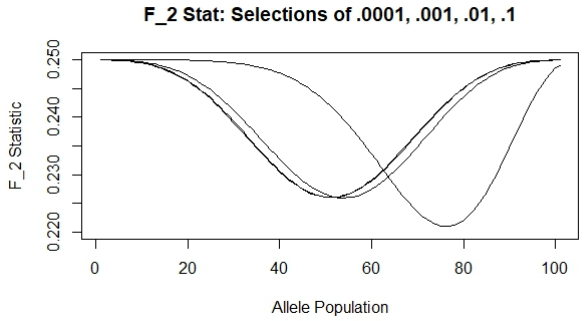


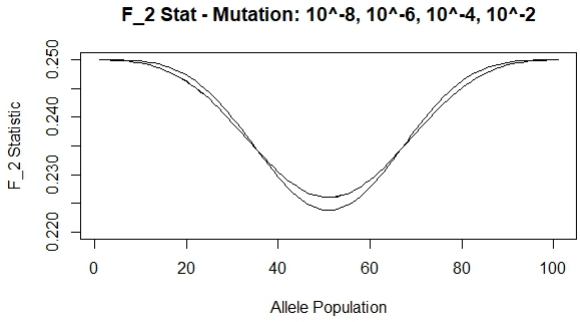
In this final set of simulations, it seems like the F_2 statistics for both groups is relatively analogous, with similar modes and distributions.

Finally, calculations were made for the PDE selection and mutation case. Default values used were 10 generations, initial allele frequency of 0.5, 10^{-4} selection coefficient, mutation rates of 10^{-8}, and a total population size of 100. The F_2 statistic was looked at under changes in initial allele frequency, generations, selection coefficients, and different mutation rates.





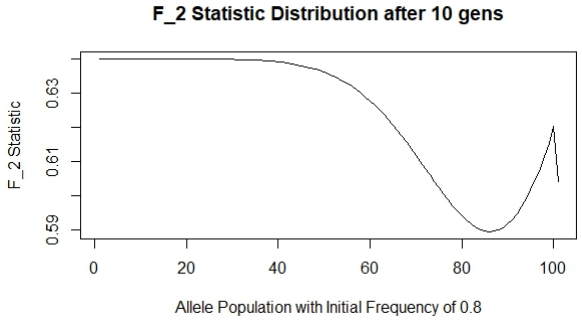




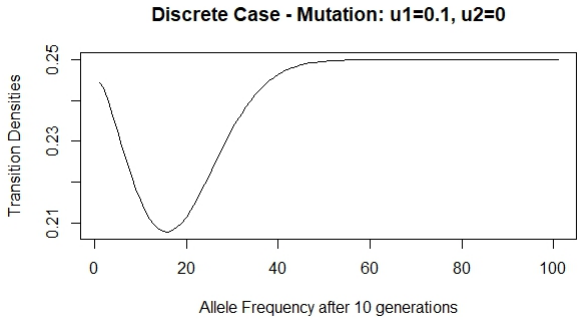
One may see that these trends are the exactly the opposite trend as those seen for the simple plotting of the PDE model with selection and mutation. Moreover, when compared to the F_2 statistic under selection, the trends were identical.

The only trend that was slightly different than those observed above was when the mutation rate of u2 was held at 0, and

the mutation rate of u1 was varied. Before, changing the mutation rate of u1 from 10^{-8} to 10^{-2} shifted the binomial distribution slightly to the left. In the selection-mutation case, the base graph (with u1=10^{-8}) looked like this:

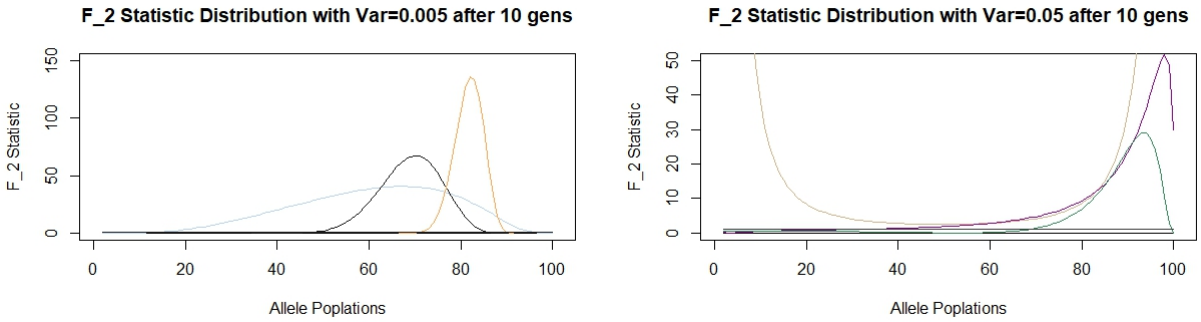


However, the u1 had to be increased until it reached 0.1 for there to be any qualitative difference in the overall F_2 statistic trend:

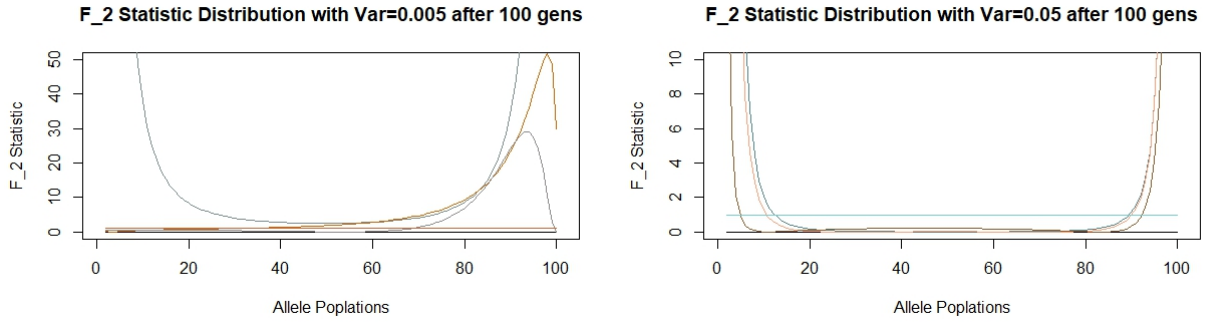


Moreover, when this change was made, there was a great qualitative change in F_2 statistic patterns.

Finally, simulations for the PDE model with randomly fluctuating selection intensities were ran. Specifically, variances of 0.05 and 0.005 were ran over 10 and 100 generations. The results can be seen below.



Again, we note that the time and variance of fluctuation in selection intensities are directly proportional. In these simulations as well, the F_2 distributions for 10 generations with a variance of 0.05 and for 100 generations with a variance of 0.005 are identical.



4 Conclusion

In this paper the author analyzed and develops the F statistic under models of selection. The author used both stochastic PDE models and discrete models with additive selection, random fluctuation of selection intensities, and selection with mutation to expand the scope of these statistics in theoretical population genetics.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors have contributed to all parts of the article. All authors read and approved the final manuscript.

References

[1] Green, R., J. Krause, A. Briggs, T. Maricic, U. Stenzel et al., 2010 A draft sequence of the Neandertal genome. Science 328: 710–722.

[2] Haak, W., I. Lazaridis, N. Patterson, N. Rohland, S. Mallick et al., 2015 Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522: 207–211.

[3] Lazaridis, I., N. Patterson, A. Mittnik, G. Renaud, S. Mallick et al., 2014 Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513: 409–413.

[4] Lipson, M., P.-R. Loh, A. Levin, D. Reich, N. Patterson et al., 2013 Efficient moment-based inference of admixture parameters and sources of gene flow. Mol. Biol. Evol. 30: 1788–1802.

[5] Patterson, N. J., P. Moorjani, Y. Luo, S. Mallick, N. Rohland et al., 2012 Ancient admixture in human history. Genetics 192: 1065–1093.

[6] Peter, B., 2016 Admixture, Population Structure, and F-Statistics. Genetics 202: 1485–1501.

[7] Reich, D., K. Thangaraj, N. Patterson, A. L. Price, and L. Singh, 2009 Reconstructing Indian population history. Nature 461: 489–494.

[8] Reich, D., N. Patterson, D. Campbell, A. Tandon, S. Mazieres et al., 2012 Reconstructing Native American population history. Nature 488: 370–374

[9] Raghavan, M., P. Skoglund, K. E. Graf, M. Metspalu, A. Albrechtsen et al., 2014 Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 505: 87–91.

[10] Stratton, J. A., P. M. Morse, L. J. Chu, and R. A. Hutner, 1941 Eliptic Cylinder and Spheroidal Wave functions. Wiley, New York.