Spontaneous mutations in population dynamics, a Fisher perspective

Arthur PLAUD

Goal The main goal of the project is to study the role of spontaneous genetic mutations inside a population. Common causes for these mutations include environmental threads such as electromagnetic radiations, exposure to chemical products but also manufactured products, where the two main exemples are drugs and cigarettes.
Fisher model is an old model, maybe the simplest one ever introduced to study genetic drift, also called natural selection, which is the way the most fit genotypes are selected inside a population.

System The precise dynamical system is a generalization of Fisher evolution model, with a stochastic term taking into account spontaneous mutations.
The time varying quantities are the $N(\omega, t)$, which describe the number of individuals of fitness ω at time t. The fitness, introduced by Fisher, is a summary of the genetic information of an individual and determines how long it will live in average, as well as how many offspring he's expected to have.
These number of individuals are related by the following equations :

$$\frac{\partial N}{\partial t} = [f(\omega) - \langle f(\omega)\rangle]N + \int_{-\infty}^{+\infty}[g(\omega, \omega', t)N(\omega', t) - g(\omega', \omega, t)N(\omega, t)]\mathrm{d}\omega'$$

In this equation, the first term describes genetic drift : individuals with higher fitness tends to live longer and have more offpsring, so their number rises inside the population. To match this description, $f(\omega)$ is taken to be a monotonically increasing function of ω. Worth noticing is that, as for the original Fisher model, this is a fixed population size model : the relative proportion of genotypes inside the population changes, but the total amount of individual stays fixed. The second term, the stochastic one, also satisfies this condition, as it's written as a classic global balance term.
This observation leads to the necessity of adjusting the model (especially the 'normalization' term $\langle f(\omega)\rangle$) to study population growth. Indeed we expect a high rate of spontaneous mutations to destabilize the population and lead to extinction.
Inside the stochastic term, the quantities $g(\omega, \omega', t)$ describe how spontaneous mutations modify the fitness of some individuals. These are stochastic quantities and we can choose different frameworks to twist the model. The most natural way of defining these stochastic transition rates is to make them independant and identically distributed in time, that is :

$\langle g(\omega, \omega^{'}, t)\rangle = h(|\omega - \omega^{'}|)$ and $\langle g(\omega, \omega^{'}, t) g(\omega^{''}, \omega^{'''}, t)\rangle = \langle g(\omega, \omega^{'}, t)\rangle \langle g(\omega^{''}, \omega^{'''}, t^{'})\rangle$ if $t \neq t^{'}$.

These hypotheses could be modified further in the study but seem like a good starting point. For exemple in real populations, mutations will typically decrease the fitness level so the symmetry in h could be relaxed.
There are a lot of ways the function h can be chosen, but there seems to be 2 interesting behaviours to study : a constant h, meaning there is no correlations between the level of fitness before and after a mutation, and a strictly decreasing h, meaning mutations with the biggest change in fitness have the lowest probabilities.

Dynamical and Computational Properties As this is a system with a stochastic description, we are not that interested in single runs but more on a statistic description of the behaviour. What is the long term behaviour of the system in term of average fitness for example. Is there any sensibility to initial conditions ? As there are many control parameters of the system (average spontaneous mutation rate, width of the curve h if we're not in the flat case, ...), the dependance of the system with these parameters will also be studied.
As an information related quantity, the entropy of the distribution $N(\omega, t)$ and its change with time will be very interesting. This entropy allows us to quantify the amount of disorder in the population : for a homogenous genotype, the entropy will be low, and it will increase when the number of relevant fitnesses inside the population will increase. This is a good order parameter. For now, I don't see a lot of connections with the study of discrete time hidden Markov chains we did in the lecture, but I will probably find some along the way.

Methods The study will be a combination of numerical simulations and analytic computations, with the analytic part mainly focusing on computing statistics of the system (stationary distribution at long time scales, average fitness, average entropy...), whereas the simulations will focus on the transient part, the sensitivity to initial conditions and fluctuations, nontypical behaviours.
For analytic calculations, I only expect the case $h = h_0$ to be solvable, but I will look for ways of including correlations in a simple enough way.

Hypothesis The main reason I engaged in this project is to study the transition from order to disorder. Indeed, in the usual Fisher evolution model, the high-fitness population is deterministically increasing and the system tends to uniform genotype. Actually, the model for two values of fitness leads to logistic growth of the higher fitness type, and disappearing of the least fit one.
Here, with the addition of spontaneous mutation, which is a form of noise in the system, we expect this order to be weakened. In the high mutation rate limit, we actually expect the stationary distribution to be uniform over the possible fitness values. The study of this transition will be the major focus.
Once I've included time-varying population size in the model, I'm also expecting that for a high enough mutation rate, the population will die out at finite time, as we know high mutation rate is typically non sustainable for real populations.

Steps and Time taken - Writing the first code to run simulations and look for interesting behaviours to study. Check that some of these properties can be studied analytically
Writing the code will be pretty simple, but finding the right parameter range, timescale to run the simulations and choosing the right order parameter could take some time. The time needed is not fixed though, as there is a part of luck involved. I would estimate it to be between 4 and 10 days.
- Once I'm settled on what I wanna show, optimizing the code to improve the statistics. I will then be able to use higher quality plots to prove my points.
Here, it's the coding part that could take time, I would probably guess 5 days.
- Actually running the long simulations and in the meanwhile doing the analytical calculations corresponding to some of the numerically-shown properties.
1 week should be enough for this, it will depend on how many things I wanna study in the system but for a 1-month project, this amount will probably be limited.
- Writing the report
Looking at my past experience in term of projectsof this size, I will estimate it to be 3 days.
This leaves me 1 week of margin for everything that will go bad, mainly numerical issues or the "looking for interesting stuff" part at first step.