Evolutionary Dynamics

Supreet Saini

Chemical Engineering

Indian Institute of Technology Bombay

Week 11

Lecture 53

Hi everyone, welcome to the next video of the course. So, we will continue our discussion of some of the interesting results that have come out from the LTEE course. Over the last three and a half decades. But for the purpose of this video, we will continue our discussion from where we left off last time. And we were faced with a conundrum as to how to view an evolutionary process happening in real time.

And the two choices that we are looking at are that if I were to plot time, which is measured in number of generations processed versus fitness of the population, then from extensive experimental evidence, I know that this trend is going to look like the following. That at t equal to 0, the ancestral population has some fitness called F naught, and as time moves forward, this rate of increase of fitness begins to slow down. It starts off with this rapid increase in fitness with time and slows down continuously. For those of you who are not clear about what I mean by that, let us imagine that in this much amount of time. At the start of the experiment, the increase in fitness that happened was that the fitness went from F naught to F1. However, if I were to take a similar-sized time window at a later point in the experiment, so these two time intervals are the same, then later on in the experiment over the same period of time, the change of fitness that takes place can be estimated as the difference between these two fitnesses, which let us call it F2 and F2.

This we can call as F3. So, the delta fitness is F1 minus F naught in the early phase and F3 minus F2 in the later phase of the experiment, and F1 minus F naught is much greater as compared to F3. As compared—yeah, at t equal to 0, this is a fitness. So, this number is much greater as compared to F3 minus F2, and hence we are saying that with the passage of time, the rate of adaptation is slowing down. This much amount of time facilitated a much larger increase in fitness at the start of the experiment.

But the same quantum of time facilitated a much slower increase in the much slower increment in the fitness as we moved forward. This has interesting implications as to what we view is going to happen as this experiment is allowed to move forward further on. And what we discussed at the end of last video was that. From here on onwards, we have two choices about what we view as the future course of this experiment. In one view, we are going to think of this as a fitness landscape.

A fitness landscape is 4 to the power L nodes. Each node has a fitness. Hence, no matter how large a number 4L is there must be one node among these 4L sequences which is actually the peak fitness beyond which you cannot go and let's say that peak fitness is represented by some number called Fmax and in this view fitness can never cross Fmax so as this experiment is allowed to move forward in this view The fitness will keep on increasing, but it will never cross Fmax.

So mathematically, this can be said as fitness will asymptotically approach Fmax. So this is the first view that fitness asymptotically approaches Fmax. this peak fitness, which is Fmax. But then the problem associated with this view, if our intuition of evolution subscribes to this view, then the problem that we have going forward is that this says that evolution will grind to a halt. Because what this is telling me is that as I keep moving forward in time, the rate of increase of fitness is going to get slower and slower with time.

And on top of that, there is a threshold that the fitness can never increase. It can never go beyond. And as a result of this, this is going to plateau here, and there is no adaptation taking place after some time. So, if we subscribe to this Fmax view, then we are also subscribing to the fact that those E. coli in the flask that Lenski and his team are evolving will reach a grinding halt where nothing really great is going to happen, which is possible. We do not know that yet.

The alternative view is that the fitness here will keep on increasing, but the rate of increase is going to get slower and slower as we move forward in time. In this view, there is no peak. There is no maxima. And fitness just keeps on increasing. The only thing is that the rate of increase of fitness is continuously slowing down.

So, the rate of increase of fitness continuously slows down. If this is the case, then Evolution doesn't grind to a halt because we are saying that it will keep on evolving. It will just keep on evolving slower and slower. But there is no maxima threshold that we are saying fitness cannot cross.

But if this view is true, then we have to ask ourselves how we view landscapes then. We've been discussing this landscape analogy for quite a while—that there are four to the power of L nodes, two nodes which are separated from each other with a distance equal to one are connected because you can acquire a mutation and move to the other node. And we raise each node by a height proportional to its fitness. And that is how we are going to view this evolutionary process, equivalent to climbing this hill. However, if there is no maxima, then how is it that 4 to the power of L points can lead me to a structure which does not have a maxima?

So there are these two conundrums associated with each of these two points. And about 10 years ago, Lenski and co-workers decided to ask this question in an experiment—I thought that was really cool. So what did they do? They did the following. This experiment had been going on for some time.

This is a study that was published in the early 2010s. So this experiment had been going on for well over 20 years. So it had lots of generations already processed. And the question they asked was: what happens if I plot time versus fitness of these evolved lines? So this experiment, let's say, has been going on for 40,000 generations.

That is the amount of data that they already have. By this time, they started work on this. Now, 40,000 generations they are processing. Remember, they are doing a dilution of 1 is 200, which is equivalent to 6 to 7 generations in every single day. Now, you won't quantify the fitness of these populations after every 6 generations.

That's simply not feasible, nor is it required. So, what they do is they quantify fitness of these evolved lines. So, if the line is starting from 0 and going to, let's say, 40,000, What they do is they quantify fitness every 500 generations. So at 500, when the evolved populations reach 500, they quantify the growth rate and that's the proxy for fitness.

They do the same for 1000 and 1500 and 2000 and so on and so forth. So they have fitness data with respect to time like we have posed here. Now, at t equal to 0, this will have some fitness. Let us say at t equal to 0, this number is F naught. And then every subsequent 500 generations, they have a data point.

And that is a data point for one line. They have similar data points for all 12 lines. And this data, let us say, looks something like this. Let's say the data looks something like this. Now, obviously, we are saying that they did this for the first 20,000 generations.

And if they are measuring every 500 generations, that means it's 40 readings. This graph clearly doesn't show 40 readings, but you get the idea. So there are these 40 data points that we have for every single line, which gives me a trajectory of how fitness changes over time. And fitness is simply the growth rate here. So, if that is the case, the next thing they do is try to fit a mathematical function to these points.

So, we have these points and fit a mathematical function. to these points to get an estimate of the trend we are looking at as these lines evolve. Of course, this is a trend that I can qualitatively see and say that it's increasing faster and then slowing down. But can I give it a mathematical form? And maybe that mathematical form allows me to distinguish between this viewpoint and that viewpoint.

And that is the goal of trying to make this quantitative in this experiment. So they do that. And interestingly, they don't fit just one mathematical formula. They try to fit two mathematical formulas. And we'll very soon see why they did that.

So in the first one, fitness or the y-axis. So let's say this is y. Fitness or y is simply equal to F naught. So, another mathematical trick we can do there is represent fitness as the y-axis as the fitness of the line divided by y naught. What that does is I am normalizing every fitness value with y naught, and what that this should be f naught. I am dividing every fitness value with f naught.

f naught, remember, is the fitness of the starting genotype. Hence, the fitness at t equal to 0 simply becomes 1 because this just becomes f naught by f naught. So, I am starting with fitness 1. So, the trend, the first trend that I will try and fit is simply given by 1 plus a times t upon b plus t. t here is the time x-axis, which is measured in number of generations. And a and b are constants.

A and B. So this is the exercise in data fitting. These are constants whose values have to be determined so that this curve can approximate the experimental data that I have to the best possible extent. Constants to be determined. And we can do some mathematical exercises, and we don't need to get into the details here. We can do some mathematical exercises and find the values of A and B. Let's call them A naught and B naught, which best describe the experimental data.

And they did that. And what we find is that by choosing appropriate values of A and B, which are A naught and B naught, when we plug in this function y as a function of t and find out what the mathematical curve tells me, we get an excellent fit with the experimental data. So, the fit of this model is great. This means that this model is able to recapitulate the experimental data via an excellent fit. It is excellent, and what they try to do is quantify how good the fit is by computing the R-squared value.

And it was well over 0.9. So, this model explains the experimental data very well. If that is the case, then if this is a good model that explains the experiment we have, then this model can be used to make predictions. This model has a unique—let us do it on the next slide. So, what we have here is experimental data.

And then we have this experimental data. Then we had this model called 1 plus AT upon B plus T. Let us call it A naught and B naught. And when I plot this model, I get an excellent fit with the experimental data. This is T measuring time, and this is Y measuring normalized fitness, where all fitness values are normalized with the ancestor's value. So, the R-squared of this fit is well over 0.9.

So, that means this is an excellent fit. So, if this is a good fit, then the question I am going to ask is: What does this tell me about the nature of the relationship between the number of generations and the fitness of a population? Now, this is just a constant. So, we do not need to worry about it. What is this term?

This is simply A naught T upon B naught plus T. When T is zero, this quantity, A naught T upon B naught plus T, is also equal to zero. And when T approaches infinity, we can think of this as A naught multiplied by a number which approaches infinity, divided by B naught plus a number which approaches infinity. B naught is negligible as compared to a number which approaches infinity. So, this whole ratio approaches A naught. What this means is that this fitness, this function

If I if I another way to look at this is that I can view this fraction as a naught into t upon b naught plus t. Now, as you can see that no matter what value of t I put here in this bracket term, the denominator will always be larger than numerator, which means this number will always be less than 1. It can get very close to 1 when t is a extremely large number, but this ratio will always be less than 1, which means that this second term will always be less than a0, which means that this mathematical expression between time and fitness tells me that fitness, this number is 1, this number is always less than a0, which means fitness or y is always less than 1 plus a0. it does not matter what is the value of T that you put in, fitness will always be less than 1 plus A naught, which means here is the asymptote that we were talking about as per the first view that fitness will never cross and this asymptote is 1 plus A naught. So, this model predicts something and it says that this is the threshold of Fmax that a population will never be able to cross. OK, so that's that's first model that they tried to fit.

And as I said that they actually fit two models. So now we look at the second one and see if that one tells us anything, something any different. So again, we go back to the sketch. And we get T versus Y with Y is the normalized fitness. And now the expression that they use is Y is equal to AT plus 1 to the power B.

So, that is the expression that they use in the second formulation of the relationship between number of generations represented by T and normalized fitness represented by Y. So, first thing we should check is that again the goal here is to find the values of constants A and B such that this model is able to capture the data as best as it can. So we can find some values of a and b and they come out to be positive. So at t equal to zero, this model just becomes one to the power b, which is simply one. So at t equal to zero, the value is always one. So I'm starting with one anyway.

So that works. And as t approaches infinity, this model tells me that y approaches a times something that approaches infinity plus 1 to the power b and that approaches infinity. So this is the model which predicts that there is no maxima associated with the fitness landscape. Fitness will keep on increasing forever, although its rate of increase will slow down. So again, I solve this, although I am not worried about at this point in the model, we are not worried about these really long times.

We only want to find the value of let's call it A dash and B dash, which will explain the experimental data for 20000 generations as best as it could. So Y is equal to A dash T plus one to the power B dash. That's the goal that we get. that is the model that we get with these precise values of a dash and b dash. So, now, if I compare the model and the trajectory that the model predicts with the experimental data that I have, I see that the fit between model and experimental data is again excellent.

So that doesn't solve my problem because the first model gave an excellent fit and this R square was well over 0.9. I think it was the order of 0.95 or something. And the same is true for this second model. So this model, the fit with experimental data is excellent again. So, this did not really help me solve the problem that I started with.

Both models are predicting the experimental data exceptionally well. R-squared in this case was also around 0.95. So both models are predicting the experimental trends with equal and very high accuracy. So using what I have done so far, this second function—let us write it down here—is 8t plus 1 to the power of b. Both of them are predicting this experimental data very, very well. So it does not really help me resolve this problem that I am faced with.

So what do I do? So the next step in the paper that they take is a clever one. What they do is they ask the following question. Remember that the experiment has happened for 40,000 generations, whereas the model was built with only data from the first 20,000 generations. So I am going to

represent the 20,000-generation data in red. So, this is Y normalized fitness, and this is T, the number of generations. So now the idea that we have ahead of us is that this point is 20,000 generations, and going forward, I have data for another 40,000 generations because that experiment has already happened. Now, I have this data for the first 20,000 generations. And using this data, I predicted the evolutionary trajectory from the two different models that I had.

So, let us say model 1 was y equals, this was the one which said y is equal to 1 plus a t upon b plus t. And let us say model 2 was the one which said y is equal to a t plus 1 to the power b. And they have different implications with regard to fitness peaks. So, let us say that model 1 both gave excellent fits. So, model 1 made a prediction like this. And model 2, let us say, made a prediction which was like this.

So, both models made predictions which were excellent in nature, with R squares well over 0.95. But now they did something clever. So, these models were fitted with data from 20,000 generations. So, they only used data from the first 20,000 generations to define the value of A naught, B naught, and A dash and B dash. So, these values are frozen because of the data that I am using to fit with the first 20,000 generations.

And now they ask which one of these two models is able to better predict what is going to happen in the next 20,000 generations. So, data is only fit using this. But these models can be extended by putting the value of t from 20,000 to 40,000. So, these models are going to make predictions, and this model 1 makes a prediction such as this. So, this curve is for model 1.

Whereas model two makes a prediction which is a little bit higher. Remember, model two is the one which says that there is no no upper limit to what is the fitness value that can be achieved. This is model two. So, again, I am emphasizing that model was frozen from the data of the first 20,000 generations. And after choosing A naught, B naught, A dash, B dash from the data generated of the first 20,000 generations, we are asking the question that which of these two models is able to better predict what is going to happen in the next 20,000 generations.

And when the data, so this is model prediction. And after that, you go and do the experiments of the strains which are in this window of time. And what the data shows us is that the data is much closer to model 2 as compared to model 1. This model turns out is a better predictor of evolutionary change. As compared to the first one.

But that leaves us in a conundrum as to what happened to the if if there is no maxima associated with fitness, because this model says that fitness can practically become infinite given infinite time. But then what does it do to my understanding of fitness peaks and the fact that fitness, there are laws of physics which will constrain the rate of divisions and so on and so forth. Those are discussions that we'll continue with in the next video. But this was a surprising result that it's actually model two, which is a better predictor. We'll continue this discussion in the next video.

Thank you.