Luria Delbrück Experiment
Experimental biophysics

Seppe Kuehn

February 25, 2020

1 Introduction

Evolution by natural selection is the organizing principle which unites all aspects of biology. It has been argued that to understand biological systems we must understand evolution - and there are still a lot of open questions about how evolution works. We're a long way from a predictive theory!

The experiment you will perform in this module recapitulates a result which stands as one of the most important contributions of quantitative thinking to all of biology. The reason is simple – the results *require* mathematical insight to correctly interpret the experiment and the question that is addressed is of profound importance to understanding the mechanisms by which life evolves. As is often the case with important contributions - the approach is simple in retrospect!

Natural selection as conceived by Wallace and Darwin in the 19th century has three main components: variation, inheritance and competition. Variation is necessary to generate differences between individuals in a population. These variations must then be passed down through generations (inheritance) and impact the growth or fitness of an organism during competition. The experiment we will perform in this module asks a clear question about precisely how variation happens.

We would like to know whether mutations in the sequence of DNA that allow an organism to adapt occur at random, and are subsequently selected during growth, or whether mutations are responsive to the environment. You can think of the difference between these hypotheses as asking whether a Neo-Darwinian picture of evolution is correct or a Larmarkian picture is correct. In a neo-darwinian picture mutations are constantly happening randomly in a population of organisms. Mutations that help the organism compete are then selected for through competition. We will call this the "**random mutation hypothesis**." The other possibility we will call the "**adaptive mutation**" hypothesis is a Lamarckian mechanism of adaptation. In this case mutations arise *after* the selection pressure is applied. In the Lamarckian mechanism mutations occur in response to selection. An interesting historical note – Darwin himself believed in Lamarckian mechanisms of variation!

2 A quantitative test

In 1943 Salvador Luria and Max Delbruück set out to ask whether the random or adaptive mutation hypthesis was correct. They devised an ingenous quantitative experiment to answer the question. Your goals are to (1) repeat this experiment and use simulations to determine whether mutations are adaptive or random, (2) use the method you develop to measure mutation rates in two strains of bacteria.

We begin with a brief overview of the experiment and a qualitative argument for how we can discriminate between the two hypotheses. We then formally treat the problem and sketch out the simulation you should perform.

2.1 Random mutations result in large fluctuations

Imagine a growing population of bacteria. In a single test tube of *Escherichia coli* there are $\sim 10^9$ cells. If you seed a test tube with say 10^3 cells and sufficient nutrients they will double roughly every half hour for about 10 hours resulting in 20 generations or a $2^{20} = 1 \times 10^6$-fold increase in the population size (Fig. 1A). At the *end* of this growth phase we could apply a selection pressure – in our case we will expose the cells to an antibiotic (rifampicin (Fig. 1B)).

Intuition: We begin by making a qualitative argument for how we expect the experiment to come out under the two hypotheses. Consider Fig. 1B – in the left panel we show two generation of bacterial growth with each cell becoming two daughters. In the adaptive hypothesis the individuals in the growing population are *susceptible* to the antibiotic (green dots). As a result, none of them are resistant to the selection pressure when it is applied. Upon application of the antibiotic a few cells (red) acquire resistance in response to the pressure and survive. This is the adaptive hypothesis.

Contrast this with the random hypothesis (Fig. 1B, right panel). Here growing populations randomly acquire mutations *before* the selection pressure is applied. In this case resistant mutants (red cells) are present before they encounter antibiotics. We call this the random hypothesis.

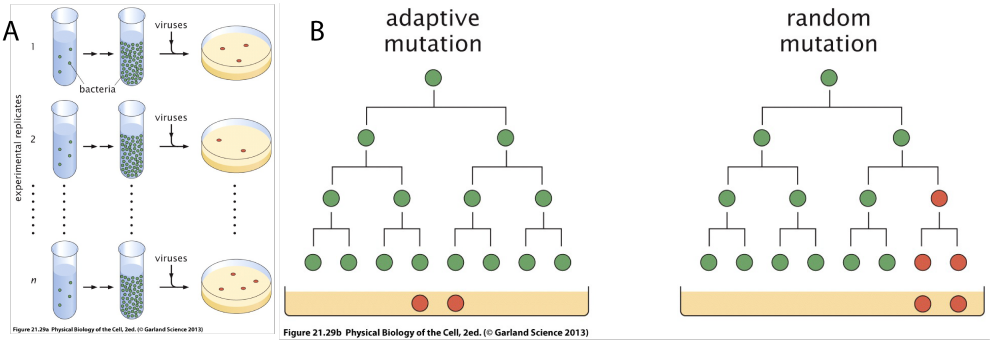

Figure 1: Growth and the adaptation vs random hypotheses. A. Consider a test tube with growing bacteria that start at some low population, grow for a period of time before being plated onto media which applies a selection pressure to the cells (virus in the case of the original paper, antibiotics in our experiment). B. Sketches for the two hypotheses. In the adaptive hypothesis no mutants arise during growth (left) but only appear *after* the selection pressure is applied (bottom). In the right panel mutants arise during growth (red cells) and are present in the population *before* the selection pressure is applied.

Under the random hypothesis if any one cell mutates and becomes resistant all of the daughter cells from that individual will also be resistant. Now imagine a population growing for many generations. If, randomly, an individual early in the growth phase acquires a mutation this will result in a HUGE subpopulation that has resistance because that individual had a lot of daughters, grand daughters...etc (we assume mutants grow at the same rate as non-mutants). Luria and Delbrück called these jackpot events. In contrast, under the adaptive hypothesis each cell has some probability of mutating after the selection pressure is applied - but in this case each cell is an independent event - whether one cell mutates is independent of another cell mutating. One way to quantify variability is to compute a Fano factor (ratio of variance to mean in the number of mutants σ^2/μ) which we will show is significantly larger under the random hypothesis.

As you will see below there is a remarkable quantitative implication of considering these two hypothesis mathematically. Namely, under the random hypothesis, once in a while there will be a muta-

tion early in the growth phase that results in a large number of resistant mutants present. Under the adaptive hypothesis the probability of observing a large number of mutants in any one population is astronomically small! What this means is that if we look at an ensemble of populations (Fig. 1A) and we consider the variability in the number of mutants we observe we will find a small variability under the adaptive hypothesis and a large variability under the random hypothesis.

2.2 Interim Summary:

Therefore, in order to test the hypothesis (random vs. adaptive) we need to quantify the variability *in the number of mutants across independent populations.* As is shown below, in the theory section, the right way to measure this variability in our experiment is the Fano factor – the ratio of the variance to the mean in the number of mutants we observe across populations. We show below, that for the adaptive hypothesis we expect the Fano factor to be 1, while for the random hypothesis we expect the Fano factor to be $>> 1$. Therefore, to test our hypothesis all we need to do is to experimentally *measure* the number of mutants that arise in a large number of independent populations (wells in a 96-well plate) and compute the average and variance in the number of mutants and take the ratio. If this number is $>> 1$ it supports the random hypothesis, otherwise, it supports the adaptive hypothesis. We want **you to test this hypothesis by computing the Fano factor on your data.**

2.3 Objectives

Questions you are expected to answer in your laboratory report:

- For wild-type *E. coli* (MG1655) perform a Luria-Delbruck experiment (see protocol) to measure the mutation rate for this strain.
- For MG1655 **compute the Fano factor (described below) for your data** and answer the question: are mutations random or adaptive?
- We will provide you with two additional strains of *E. coli* with higher mutation rates than wild type (mutator strains, genetic variants of *E. coli* with a higher mutation rate). For at least one of the mutator strains perform another Luria-Delbruck experiment and measure the mutation rate in this strain. How does it compare to what you measured above for the wild-type?

Theoretically (and computationally) we would like you to complete the following objectives. These are explained in more detail at the end of the document:

- Convince yourself that we can measure the mutation rate by counting the number of populations we observe with zero mutants (Eqn. 10).
- Follow the derivation of the Fano factors for the two hypotheses.
- (Additional exercise 1) Perform a numerical simulation, using Matlab, to compute the Fano factor under the two hypotheses. Use your measured mutation rate. How do the Fano factors compare? Use the adaptive hypothesis as your null model and compute a p-value for the experimentally observed Fano factor (for MG1655) under the null that the adaptive hypothesis is correct.

- (Additional exercise 2) How does the Fano factor you computed for MG1655 compare to your analytical expectation? For example, if your data support the random hypothesis, is there quantitative agreement between the Fano factor you get from your simulations or is one much different from the other? Why might these numbers differ? One reason could be that mutants grow more slowly than susceptible bacteria! Modify your simulation from exercise 1 to include this effect - how does it change the Fano factor you observe in simulation?

- (bonus) Explore the meaning of mutation rates using stochastic simulations of genomes as described in Additional exercise 3.

3 Theory

3.1 The adaptive hypothesis

Let's start by considering a population of N cells that are exposed to a selection pressure. Say there is a probability a that an individual adaptively mutates to resist the antibiotic. We now compute the probability of observing m individuals ($m < N$) that are resistant. The probability that m cells mutate is just a^m. We need to multiply this by the probability that $N-m$ cells do not mutate ($(1-a)^{N-m}$). The cells themselves are not labeled in any meaningful way so we need to account for the number of possible ways of choosing m mutants from N cells and the result is the binomial distribution:

$$P(m,N) = \frac{N!}{m!(N-m)!}a^m(1-a)^{N-m} \tag{1}$$

a is the mutation rate – just the chance that a cell successfully mutates. It turns out that $a << 1$ (e.g. 1×10^{-9} or so). Also, it is intuitive that the average number of mutants you should get if you performed the experiment many times is $\langle m \rangle = aN$. This is because you have N individuals each with a chance a of mutating. Since a«1, on average m«N. Using this limit we can make the following simplifications.Note, in this document's text, we will use () to denote multiplication and [] to denote function. Some figures do not make this distinction.

$$P[m,N] = \frac{N(N-1)..(N-m)}{m!}a^m(1-a)^{N-m} \tag{2}$$

$$P[m,N] \approx \frac{N^m}{m!}a^m(1-a)^N \tag{3}$$

$$P[m,N] \approx \frac{(aN)^m}{m!}(1-a)^N \tag{4}$$

Now recall the following limit:

$$e^x = \lim_{N\to\infty}\left[1+\frac{x}{N}\right]^N$$

Now rewrite Eqn. 4 as

$$P[m,N] \approx \frac{(aN)^m}{m!}\left(1-\frac{aN}{N}\right)^N \tag{5}$$

$$P[m,N] \approx \frac{(aN)^m}{m!}e^{-aN} \tag{6}$$

and use the average we mentioned above to find the result that:

$$P[m, N] \approx \frac{\langle m \rangle^m}{m!} e^{-\langle m \rangle} \tag{7}$$

Equation 7 is the Poisson distribution, and it is true anytime N is large and the probability a is small – anytime you have a binomial distribution in this regime the limit derived is correct.

Optional computational exercise: *Check that a Poisson distribution is a good approximation to a binomial when N is large and a is small. At values of N and a does the approximation start to break down?*

Equation 7 let's us compute how many resistant cells we expect to find in a population of N cells if we assume that the adaptive hypothesis is correct. That's great, but to do this we need to know two things: (1) the number of cells N – which we can measure (see below) and (2) the mutation rate a. What is a? How do we compute it?

3.2 Measuring the mutation rate

Luria and Delbruück presented a remarkably clever way to compute the mutation rate which works irrespective of which of the two hypotheses is correct! Imagine there are N cells – this means that there were N chances (division events) to get a mutation that conferred resistance to the antibiotic. Under the adaptive hypothesis the Poisson distribution tells us how many cells will be resistant, but under the random hypothesis it does not! Why? Because if a cell mutated early during the growth phase it would give rise to many many daughter cells that are resistant because they inherited the mutation (Fig. 1B) but didn't mutate themselves. However, the key insight is that the Poisson distribution correctly describes the probability of observing zero mutants *in both cases!* The reason is that under both hypotheses if no mutations occurr during growth there will be no mutants! (We assume that there were no mutants originally.) Let's go back to the binomial distribution (Eqn. 1) and put in $m = 0$

$$P[0, N] = \frac{N!}{N!} a^0 (1 - a)^N \tag{8}$$

$$P[0, N] \approx e^{-aN} \tag{9}$$

We can then solve for a and get:

$$a = -\frac{ln[P[0, N]]}{N} \tag{10}$$

So we now need an estimate of $P[0, N]$ – the key insight is to realize that we can compute this by performing many parallel growth experiments (Fig. 1A) and compute the fraction of populations where we observe no mutants. Using this estimate of the probability of observing zero mutants, we can compute a directly from data! Eighty years after the work of Luria and Delbrück this remains a standard laboratory technique for estimating mutation rates.

Note that we can now compute the expected number of mutants under the adaptive hypthesis as $\langle m \rangle = aN$. We'll need this in the next section:

3.3 Mean and variance under adaptive hypothesis

Now we are prepared to compute the Fano factor under the adaptive hypothesis. To do this we need simply to compute the mean and variance for the Poisson distribution. We leave this as an exercise to the reader, recall that the mean and variance are computed as follows:

$$\mu = \sum_{m=0}^{\infty} mP[m, N] \tag{11}$$

$$\sigma^2 = \langle m^2 \rangle - \langle m \rangle^2 = \sum_{m=0}^{\infty} m^2 P[m, N] - \mu^2 \tag{12}$$

We leave it as exercise to the student to show that:

$$\sigma^2 = \mu \tag{13}$$

Which means that the Fano factor for the adaptive hypothesis is unity!

$$\boxed{F^{adapt} = \frac{\sigma^2}{\mu} = 1} \tag{14}$$

What does this mean? If we perform the experiment many times (Fig. 1) we can compute the number of mutations that are present in each of a large number of replicate populations of size N. We can then directly compute F from data and check – is it close to one or not?

Our next claim is that F^{rand}, the Fano factor under the random hypothesis, is much larger than unity. This proof requires a bit more effort.

3.4 The random hypothesis

We would now like to compute the same quantity (F) assuming that the random hypothesis is correct. The problem is that resistant mutants arise via two processes in the random hypothesis: mutations of susceptible cells and growth of mutants. Since we cannot predict precisely when during growth a mutation will occur, we need to use some more sophisticated machinery to get there. We will use a master equation formalism. A master equation is similar to writing a derivative, only it is for probabilities instead of deterministic equations like in calculus. Basically, a master equation says that the rate of change (with respect to time) of being in a state s at time t is the probability of staying in or transitioning into state s minus the probability of transitioning out of state s. For example, lets have a system with different possible states say $s-1$, s, $s+1$, etc., where there is a transition rate of moving from one state to a different state such as $s-1 \rightarrow s$. Here the probability of transitioning into state s is the probability of being in another state, say $s-1$, at time $t-1$ times the transition rate per unit time from $s-1 \rightarrow s$. Similarly, the probability of transitioning out of state s is the probability of being in state s at time $t-1$ times the transition rate of leaving state s. These concepts will be described a bit more later while introducing our master equation, but if you need some better intuition for master equations you can look up online resources such as this one (`http://biotheory.phys.cwru.edu/phys320/phys320_master_equation_I.pdf`).

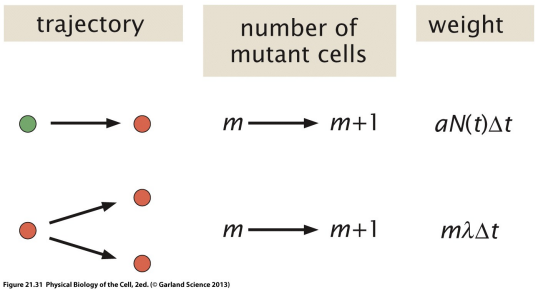

Figure 2: Generating mutants under the random hypothesis. There are two routes to generating mutants under the random hypothesis. (top) A susceptible cell mutates giving rise to a mutant. This increases the number of mutants by 1 and has a weight of $aN[t]dt$ since there are N cells each with the same chance of mutating to become resistant. (bottom) a second possibility is a mutant cell that divides in two with growth rate λ and this increments the number of mutants up by 1 as well and has weight $m\lambda dt$.

3.4.1 Expected number of mutants under random hypothesis

We want an equation of motion for the number of mutants we expect to be in the population at a given time. To do this we need to first enumerate the ways in which mutants can be generated (Fig. 2), i.e. a mutant cell doubles or a regular cell mutates. We also need to write down the transition probabilities per unit time (see right column of Fig. 2). The master equation formalism is very simple. Given known transition probabilities (rates), what equation describes the dynamics of the number of mutants in time? To answer this consider the following, remember () denotes multiplication and [] denotes a function:

$$\frac{\partial p[m,t]}{\partial t} = \text{probability of transitioning into state m} - \text{probability of transitioning out of state m} \tag{15}$$

$$\begin{aligned}\frac{\partial p[m,t]}{\partial t} &= +(m-1) \text{ mutants at } t-1 \text{ and one doubled} \\ &+ (m-1) \text{ mutants at } t-1 \text{ and new mutant arose} \\ &- m \text{ mutants at } t-1 \text{ and one grew} \\ &- m \text{ mutants at } t-1 \text{ and a new one mutated.}\end{aligned}$$

Mathematically we get:

$$\frac{\partial p[m,t]}{\partial t} = \lambda(m-1)p[m-1,t] + aN[t]p[m-1,t] - \lambda m p[m,t] - aN[t]p[m,t] \tag{16}$$

Remember our goal is to compute the Fano factor and compare its magnitude to the adaptive hypothesis. To do this we need to compute the average number of mutants after some time and the variance in the number of mutants. To proceed we compute the first and second moments of $p(m,t)$ by multiplying both sides by m and summing as follows:

$$\sum_{m=1}^{\infty} \frac{\partial m p[m,t]}{\partial t} = \frac{\partial \langle m \rangle}{\partial t} = \sum_{m=1}^{\infty} m(\lambda(m-1)p[m-1,t] - \lambda m p[m,t] + aN[t]p[m-1,t] - aN[t]p[m,t]) \tag{17}$$

Which is an admittedly messy differential equation for the first moment. We deploy some tricks to solve this equation. The first term on the right hand side yields to the following manipulation.

$$\lambda \sum_{m=1}^{\infty} m(m-1)p[m-1,t] \tag{18}$$

Make the following change of variables $m-1 = m'$ to arrive at

$$\lambda \sum_{m'=-1}^{\infty} (m'+1)(m')p[m',t] \tag{19}$$

but having less than zero mutants is not possible so we can change the limit of the sum to 0 and expand to get:

$$\lambda \sum_{m'=0}^{\infty} (m'^2 + m')p[m',t] = \lambda(\langle m^2 \rangle + \langle m \rangle) \tag{20}$$

For the third term on the RHS of Eqn. 17 we can apply the very same trick after some additional trickery:

$$\sum_{m=1}^{\infty} mp[m-1,t] = \sum_{m=1}^{\infty}(m-1+1)p[m-1,t] = \sum_{m=1}^{\infty}(m'+1)p[m',t] = (\langle m\rangle + 1) \tag{21}$$

We can now write Eqn. 17 as:

$$\frac{\partial\langle m\rangle}{\partial t} = \lambda(\langle m^2\rangle + \langle m\rangle) - \lambda\langle m^2\rangle + aN[t](\langle m\rangle + 1) - aN[t]\langle m\rangle \tag{22}$$

which simplifies to:

$$\frac{\partial\langle m\rangle}{\partial t} = \lambda\langle m\rangle + aN[t] \tag{23}$$

This is a remarkably simple result that makes good intuitive sense! The first term on the RHS is the exponential growth of mutants that are present in the population. The second term is the increase in the number of mutants due to additional mutations. Compare this for a moment to the exponential growth of a population of cells:

$$\frac{dN[t]x}{dt} = \lambda N[t] \tag{24}$$

which has the solution: $N[t] = N[0]e^{\lambda t}$ where $N[0]$ is the number of cells at time zero and must be specified. Compare Eqn. 24 to Eqn. 23 to realize that the mutant population actually grows *faster* than the non-mutant population because of the two processes involved (Fig. 2). Of course we are assuming that λ does not change for the mutants which is a reasonable approximation.

Plugging $N[t] = N[0]e^{\lambda t}$ into Eqn. 23 we get a differential equation for the expected number of mutants at time t which we can solve analytically (or just look it up) from which we arrive at:

$$\boxed{\langle m[t]\rangle = atN[t]} \tag{25}$$

3.4.2 Expected variance in the number mutants under random hypothesis

To compute the Fano factor we need to compute the variance in the number of mutants we expect. Recall $\sigma^2 = \langle m^2\rangle - \langle m\rangle^2$, meaning we need to calculate $\langle m^2\rangle$ and we're done! To do this we take the same approach, but we multiple Eqn. 17 by m^2 and applying precisely the same tricks we applied above. Doing so yields the following differential equation (check for yourself!):

$$\frac{\partial\langle m^2\rangle}{\partial t} = 2\lambda\langle m^2\rangle + \lambda\langle m\rangle + 2aN[t]\langle m\rangle + aN[t] \tag{26}$$

Again, plug in $N[t] = N[0]e^{\lambda t}$ and simplify. Then solve the differential equation or lookup the solution to get:

$$\langle m(t)^2\rangle = \frac{aN[0]e^{\lambda t}}{\lambda}(2e^{\lambda t} - t\lambda - 2) + \langle m[t]\rangle^2 \tag{27}$$

which means that

$$\sigma^2 = \frac{aN[0]e^{\lambda t}}{\lambda}(2e^{\lambda t} - t\lambda - 2) \tag{28}$$

Now, we make some approximations, namely that $e^{2t} >> e^t$ for large t and we arrive at

$$\sigma^2 \approx \frac{2aN[0]e^{2\lambda t}}{\lambda}. \tag{29}$$

From equation 25 we have that $N(0)e^{\lambda t} = \langle m[t] \rangle / (at)$ which results in

$$\sigma^2 \approx \frac{2\langle m \rangle e^{\lambda t}}{\lambda t}. \tag{30}$$

Which means that the Fano factor under the random hypothesis is given by:

$$\boxed{F^{rand} \approx \frac{2e^{\lambda t}}{\lambda t}}. \tag{31}$$

This was our objective! How does this number compare to $F^{adapt} = 1$? To compute this we need to know λt which is the number of generations that transpire during the growth of the culture. As we mentioned above, *E. coli* will grow from a population of about 1×10^3 or 1×10^4 to 1×10^9 in a test tube overnight. This means that there are $\approx 1 \times 10^9$ new cells which divide over the growth phase. How many times must the population double for this to happen? If $N[0] = 1 \times 10^4$ then the population expands 1×10^5 times and this requires $log_2[1 \times 10^5] = 16.6$ doublings of the population. This means that $\lambda t = 16.6$ which gives a Fano factor of $F^{rand} \approx 1 \times 10^6 >> 1$!

What does this mean for the experiment? This means that if we perform the experiment shown in Fig. 1A with many parallel populations undergoing many divisions and then we assay all cells for their susceptibility to antibiotics and we compute the Fano factor across replicate cultures we expect that this number should be unity if the adaptive hypothesis is correct and much bigger than unity if the random hypothesis is correct!

If you would like to read the sources on which this document is based see the original paper from Luria and Delbrück (Genetics, 28, 6, 491-511 (1943)). The derivations below are taken from the second edition of the textbook: *The physical biology of the cell* by Phillips, Kondev, Theriot and Garcia. Note that in that text there are several typos in the master equation derivation which have been corrected in the presentation below. You can also read the relevant section in Philip Nelson's text *Physical models of living systems*.

3.5 A computer simulation

In addition to the experiments explained in the experimental protocols, we would like you to perform numerical simulations of the adaptive and random hypotheses. To do this you need to write Matlab code to simulate the growth and mutations under the two hypotheses.

To accomplish this I suggest you NOT try and simulate every cell in a population that reaches a size of 1×10^9! Your computer will run out of memory quickly. Instead simulate two populations mutants and susceptible cells (which we will call wild type cells). We will walk you through the easier of the two simulations – the adaptive hypothesis.

Start by computing a from your data – it will be a number of order 1×10^{-9}. The adaptive hypothesis states that this mutation rate is the rate at which cells acquire mutations after the selection pressure is applied. When you plate cells after growth, you take a population of $\sim 10^9$ cells and plate them on antibiotics. When you count colonies, you are measuring the number of cells that are resistant to the antibiotics. Under the adaptive hypothesis, this number is Poisson distributed with mean aN. To simulate this process you simply need to take random draws from a Poisson distribution with mean aN for each "plate". Then you need to compute a variance and mean in the number of mutants that arise across all plates in a numerical experiment. Estimate a Fano factor from this.

For the random hypothesis, the simulation is more involved. In this case you need to simulate the processes of growth and mutation. To do this start a population of say 10^3 cells and use each cycle of a for-loop to allow them to double. During each doubling, generate aN new mutants from the wild-type population using the Poisson distribution mentioned above. Be sure to allow those mutants to double with each cycle.

(Additional exercise 1) Computationally address the following question: what is the probability of observing a Fano factor as big or bigger than the one you measured in your experiment (with wild-type) *under the adaptive hypothesis*? This probability is called a p-value for the adaptive hypothesis as the null hypothesis. Do compute this number run many (>1000) simulations under the adaptive hypothesis. Then ask yourself: if I were to choose one of these 1000 simulation results at random, what would the chance be that the Fano factor for that simulation exceeds the factor I measured experimentally? How do you compute this probability? What is it?

(Additional exercise 2) Mutations often change the growth rate of cells. For example, there are often 'trade-offs' where mutations that give cells resistance to antibiotics slow their growth in the absence of antibiotics (for an example of this see recent work from my lab: Fraebel *et al.* eLife, 2017). Repeat your simulations above with different growth rates for the mutants. Your objective is to answer the question: how does the Fano factor depend on how much slower resistant mutants grow than the wild-type strain? To anwer this question repeat your simulation for a set of mutant growth rates varying from 0.1% to 10% reduction in growth rate relative to the wild type. Construct a plot of the expected value of the Fano factor as a function of mutant growth rates.

(Additional exercise 3 – bonus) What does the mutation rate you infer above actually mean? Here you explore this question computationally. Your simulations for exercises 1 and 2 should have been continuous simulations of a model described by differential equations. Here we ask that you extend these simulations to a stochastic model which simulates not only individual replication events but also the genomes of each cell as it divides. You will have to keep populations in these simulations smaller in order to maintain computational tractability. We suggest your simulation has the following structure:

- Initiate your population with a single cell that is described by a genome which is a binary vector of modest length (100 'base pairs').
- In each round of the simulation replicate each individual in the population. For each base pair in the genome include a *per site mutation rate* which is the probability of flipping that site. We suggest you simulate between 8 and 10 doublings (e.g. up to 1024 individuals). Divisions can be synchronous to keep the simulation simple.
- At the end of the simulation compare the starting genome of the founding cell to the genome of each individual in the population. How many mutants are there? (where a mutant is a cell that differs at any ONE position in the genome). Given the per base pair mutation rate and the number of replications in your simulation - does the number you obtain make sense?
- For evolution to 'work' random mutations have to occur in the correction position on the genome. Pick a particular position and compute the probability of *that* position being mutated in any given individual in the population. What is that probability? What do you expect it to be?

- From your simulation where a particular mutation is the target perform a version of the Luria-Delbruck simulation *in silico* – with many replicate populations being 'plated' on an antibiotic. Only mutants with a specific mutation survive. Compute the parameter a as you would in the experiment. How does the parameter compare to the per site mutation rate you put into your simulation? Note you will have to tune your parameters so that there is a reasonable probability of the mutant arising in the population.
- What happens to a if multiple mutations are sufficient to survive the antibiotic?