Arnold Sommerfeld Center for Theoretical Physics
Ludwig-Maximilians-Universität München

Prof. Erwin Frey
B. Nettuno, T. Roth, D. Toffenetti

Nonequilibrium Field Theories and Stochastic Dynamics
Sheet 2

Exercise 6 – The Luria-Delbrück Experiment

The classical experiment of Luria and Delbrück deals with identifying the mechanism of bacterial resistances to viruses.[1] Several colonies of E. coli bacteria are grown and exposed to a virus (a bacteriophage), which normally kills the entire colony. However, there is a number of surviving cells in each colony, which indicates that some cells are resistant to the virus. At the time, there were several different hypotheses explaining the observation of resistant bacteria. Luria and Delbrück tried to falsify (or verify) the following two hypotheses, based on their experimental results [2]:

(i) *Acquired mutation hypothesis:* Bacteria can develop resistance against bacteriophage as soon as they get in contact with the virus. Hence, they actively adapt to an upcoming threat.

(ii) *Random mutation hypothesis:* Bacteria randomly mutate and develop resistances against bacteriophages during their lifetime. A contact to the virus is not needed for the process, as resistances emerge spontaneously.

In either hypothesis, the bacteria pass on the resistance to their offspring. This also means that the resistance must arise at some point during the experiment and cannot be present from the beginning, otherwise all bacteria would survive the virus.

Since it was (and still is) difficult to make experiments that directly observe the moment when resistance emergences in individual cells, Luria and Delbrück had to relate both hypotheses to macroscopically observable quantities. In this exercise, we want to re-trace their work and derive theoretical predictions of statistical quantities, which we can then compare to experimental data. This will allow us to falsify one of the above hypotheses and give a substantial prediction for the underlying mechanism of the development of bacterial resistance. The experimental setup which we need to understand to formulate our theory is described in Fig. 1 in a slightly modified way. The full setup is described in their paper.[3]

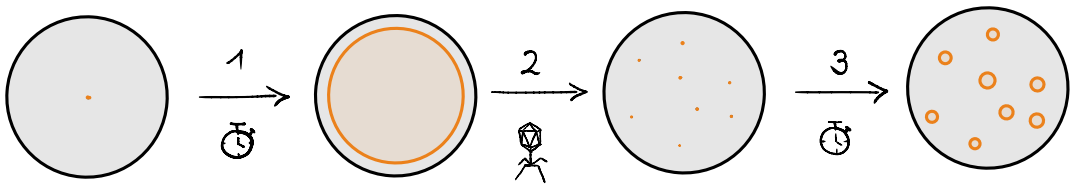

Figure 1: Cartoon version of the Luria-Delbrück experiment. It involves three main steps that are repeated many times to gather statistics: **1)** Starting from a single cell (or at least a very small number), a colony is grown to about 10^9 cells. The E. coli cells/colonies are depicted as orange dots/circles inside a petri dish. **2)** The colony is exposed to the bacteriophage which kills almost all the cells. **3)** After a shorter time, one can observe that small new colonies form around the surviving cells, which are resistant to the virus. The number of these macroscopically visible colonies is a good estimate for the number of surviving cells M. *Note that there are some details of the experiment that are ignored in this analysis. You are welcome to discuss them with your tutors (e. g.the surviving cells are really counted by diluting smaller samples and plating them onto separate dishes).*

[1] For the insights provided by this experiment, among others, Luria and Delbrück (together with Hershey) were awarded the Nobel Prize in Medicine in 1969. To learn more on Max Delbrück, see https://www.mdc-berlin.de/news/news/who-was-max-delbruck

[2] The two hypotheses are reminiscent of the Lamarckian and Darwinian theories of evolution, respectively. Although Darwinian evolution was widely accepted by the 1940's, it was not established whether it applies to microorganisms in the same way as to other animals.

[3] S. E. Luria and M. Delbrück, Genetics, 28(6), 1943, 491–511, doi: `https://doi.org/10.1093/genetics/28.6.491`.

a) *The acquired immunity hypothesis*

First, we look at the hypothesis of acquired immunity. It assumes that a mutation to resistant cells can occur as soon as the population is brought into contact with the bacteriophage, i. e., at step 2 of the cartoon in Fig. 1. In this setting, each individual cell mutates independently with constant probability p.

- Derive the probability distribution of M, i. e.the probability of generating $m \in \{0, \ldots, N\}$ mutants in a total population of N bacteria, after exposing it to the virus. How is this distribution called?
  Hint: What is the exact distribution for the probability of m "successes" in N independent, identically distributed events?
- Use $N \gg 1$, $p \ll 1$, and $N!/(N-m)! \approx N^m$ (a variant of Stirling's formula) to approximate the exact distribution by a distribution for rare events.
- Calculate the mean $\langle M \rangle$ and the variance $\langle M^2 \rangle - \langle M \rangle^2$ of this approximate distribution.

b) To assess whether the first hypothesis describes the outcome of the Luria-Delbrück experiment, we look at the experimental data from their paper.[4] In Fig. 2 we plot the variance per sample against the mean per sample for all data points (orange markers) and compare it to the prediction from the acquired mutation hypothesis (blue line). What conclusions can you draw?

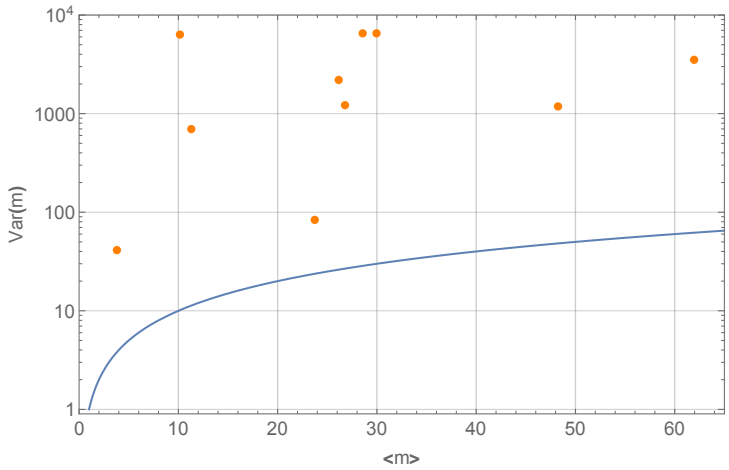

Figure 2: Variance per sample vs.mean per sample for the Luria-Delbrück-Experiment (orange markers) and the predicted relation between them from the acquired mutation hypothesis (blue line). "Per sample" refers to individual cell cultures that were prepared in the same way (step 1 in Fig. 1). From each culture, a small sample was then exposed to the virus, and the surviving cells were counted (steps 2/3 in Fig. 1). Thus, each datapoint corresponds to the mean-variance pair computed from several such cultures. Note the logarithmic scale on the y-axis.

c) *The random mutation hypothesis*

We want to repeat the same analysis for the second hypothesis. Previously, we could ignore time in the description of the experiment (only the number of cells that were exposed to the virus and the "immunization probability" a mattered). Here, in the random mutation hypothesis, we assume the following: Cells divide with a constant doubling rate λ_2 (alternatively, we could use the doubling time $\lambda_2 = 1/\tau_2$ or the natural rate $\lambda = \lambda_2 \ln 2$ to describe the process – here we will use λ for convenience). Under this assumption, their total number can be described by $N(t) = N_0 2^{\lambda_2 t} = N_0 e^{\lambda t}$. The crucial difference to before is that *during its lifetime*, every cell can randomly mutate with some small probability, which we call a, and gain resistance against the virus. The random mutations would therefore happen in step 1 of the cartoon in Fig. 1, without the cells ever being exposed to the virus.

As before, we want to arrive at an approximate formula for the variance of M (the number of mutants in a colony) in terms of its mean.

- Set up the master equation for the probability $P(m, t)$ of finding a total of $m \in \{0, \ldots, N(t)\}$ mutant cells at time t. Describe the various contributions.

[4]The original experimental data (number of surviving cells M) is given in Tables 2 and 3 of the original paper by Luria and Delbrück: `https://doi.org/10.1093/genetics/28.6.491`.

Hint: We neglect the probability of a random mutation to "undo" the newly gained resistance. Such events were practically never observed in the experiment.

- Show that for small mutation rates ($a \ll 1$), large colony numbers ($N(t) \gg 1$), and a small fraction of mutants ($m/N(t) \ll 1$), the master equation can be simplified to

$$\partial_t P(m,t) = \lambda(m-1)P(m-1,t) - \lambda m P(m,t) + aN(t)P(m-1,t) - aN(t)P(m,t)\ . \tag{1}$$

  Why/when does it make sense to assume $m \ll N(t)$, even though m could be any value in $\{0,\dots,N(t)\}$?

d) Calculate the average number of mutants at a given time $\langle m\rangle(t)$ by deriving and solving a differential equation for $\langle m\rangle(t)$ from the master equation, Eq. (1). As initial condition, we assume that there are no mutants present at the start of the experiment, i. e., $P(m,0) = \delta_{m,0}$.

e) Derive another differential equation for the second moment $\langle M^2\rangle(t)$. You don't need to solve the equation – the result is stated in the following subproblem. What is a reasonable initial condition for a solution of this differential equation?

f) The solution for the previously derived differential equation is

$$\langle M^2\rangle(t) = \frac{aN_0 e^{\lambda t}}{\lambda}\left(2e^{\lambda t} - \lambda t - 2 + ae^{\lambda t}\lambda N_0 t^2\right)\ . \tag{2}$$

Using this and your solution for $\langle M\rangle(t)$ from part **d)**, find an expression for the variance as a function of $\langle M\rangle(t)$. What is the difference compared to the acquired immunity hypothesis? How do you interpret the result with respect to the experimental results shown in Fig. 2?

Hint: It is helpful to visualize the two different hypotheses in a tree-like structure, showing one initial cell and all its offspring. Where do the mutations occur in both cases and why does one have much higher variance?

Exercise 7 – Linear Birth-Death Process

In this exercise we will consider a simple model for population dynamics which only include birth and death processes modelled as randomly occurring events:

$$A \xrightarrow{\mu} \emptyset, \qquad\qquad A \xrightarrow{\lambda} A + A. \tag{3}$$

Here μ and λ are phenomenological parameters which define the probability per unit of time (rate) for an individual to die or to reproduce. In other words, within a short time interval $(t, t+dt)$ each individual A dies with probability μdt and reproduces with probability λdt. We assume the events involving distinct individuals to be independent. To make it more concrete, imagine that A represents some species of bacteria. Depending on the environmental conditions our phenomenological parameters μ and λ can take different values: if the nutrients become scarce the reproduction rate λ may go down, if the bacteria get exposed to antibiotics the death rate μ may go up, etc. Such changes can qualitatively affect the dynamics of the population. Given the rates μ and λ, and the initial population size n_0, we would like to be able to predict the fate and the dynamics of the population, but, as we shall see, this can't be done due to the stochasticity of the dynamics. What we will do instead is try to learn as much as possible about the probabilities and statistics of possible outcomes of this birth-death process.

a) *Mean field approach*

To make first predictions for the dynamics, using minimal effort, let us start with a mean field description. Let $N(t)$ denote the number of individuals at time t. Try to write a simple deterministic equation that would best describe the average dynamics of $N(t)$. If at time t there are on average $\langle N(t)\rangle$ individuals, how do you think the number of individuals typically changes at that time? (No need to be rigorous here, you will do it properly later.) What behaviour do you expect for different values of μ and λ?

b) *Master equation*

Now we will try a more rigorous approach. Let $p_n(t) = \mathrm{Prob}\{N(t) = n\}$ be the probability of finding n individuals at time t. Derive a master equation for $p_n(t)$ and give an interpretation for all the terms in your final result, which should read

$$\partial_t p_n(t) = \lambda(n-1)p_{n-1}(t) + \mu(n+1)p_{n+1}(t) - (\mu+\lambda)np_n(t). \tag{4}$$

c) *Calculating moments*

Using the master equation find an equation for the time evolution of the average number of individuals $\langle N(t)\rangle$. How does this equation compare to the one you found in part **b)**? Solve the found equation for $\langle N(t)\rangle$ and comment on the extinction probability for $\lambda < \mu$, $\lambda = \mu$ and $\lambda > \mu$. *Hint: The extinction probability is related to the average number of individuals at long time, i.e.* $\langle N(t=\infty)\rangle$

The expectation value of $N(t)$ can give us some crude information about the dynamics of the studied population, but it may not be enough to calculate the extinction probability. In the rest of this exercise we will get a more detailed insight into the statistics of this population dynamics, by calculating the probability distribution $p_n(t)$.

d) A straightforward way to find $p_n(t)$ would be simply to solve the master equation you derived, which is in fact a system of ordinary linear differential equations. Such equations are in principle straightforward to solve, but not if there are infinitely many of them, as in our case. We will solve this problem by encoding all the functions $p_n(t)$ in a single function constructed as:

$$G(z,t) = \sum_{n=0}^{\infty} p_n(t) z^n. \tag{5}$$

The function $G(z,t)$ is termed "probability generating function". Use the master equation to find a partial differential equation for $G(z,t)$. Your result should have the form

$$\frac{\partial G}{\partial z} f(z) + \frac{\partial G}{\partial t} = 0, \tag{6}$$

where $f(z)$ is a function that you need to find. Assuming that you know $G(z,t)$, how could you calculate $p_n(t)$?

e) We managed to translate the problem of solving an infinite set of ordinary differential equations to the problem of solving a single partial differential equation of two variables z and t. This one we will solve using the method of characteristics. The idea is to find so called characteristic curves $(z(t),t)$ in the (z,t) plane on which the function G is constant, that is the level sets of G. Note that choosing a curve $z(t)$ defined by $dz(t)/dt = f(z(t))$, we can rewrite Eq. (6) as

$$\frac{\partial G}{\partial z}\frac{dz(t)}{dt} + \frac{\partial G}{\partial t} = \frac{d}{dt} G(z(t),t) = 0, \tag{7}$$

which is precisely the condition for $(z(t),t)$ being a level set of G. Then, to find the value of $G(z,t)$ for any point (z,t) we simply find the characteristic curve passing through that point and look for the intersection of that curve with the axis $t=0$. On this axis the values of the generating function $G(z,0)$ are simply defined by the initial condition $N(0) = n_0$. Therefore, we ultimately reduced the whole problem to solving a single ordinary differential equation $dz(t)/dt = f(z(t))$. This is a very special example of a much more general and powerful method of characteristics. *Hint: On the way you may need to integrate some rational function. You may find the following decomposition useful:*

$$\frac{1}{(\mu-\lambda z)(1-z)} = \frac{1}{\lambda-\mu}\left(\frac{\lambda}{\mu-\lambda z} - \frac{1}{1-z}\right).$$

f) Now is the time to reap the fruits of your labour!

- Calculate $p_0(t)$ and the probability that the population ultimately becomes extinct. You should find:

$$p_0(t) = \left(\frac{\mu - \mu e^{(\lambda-\mu)t}}{\mu - \lambda e^{(\lambda-\mu)t}}\right)^{n_0}. \tag{8}$$

- Sketch $p_0(t)$ for $\lambda < \mu$ and for $\lambda > \mu$.
- What is the form of $p_0(t)$ when we exclude reproduction? Can you give a simple interpretation of the terms in that expression?

g) How to calculate the probability distribution for the extinction time? Sketch the approximate shape of this distribution. (You don't have to give an explicit formula.) Could you in principle do the same thing to calculate the probability distribution of the time when the population reaches the size of 100 individuals for the first time?

Exercise 8 – Linear Birth-Death Process – Numerics

After having calculated some analytical results for the linear Birth-Death Process, we now want to compare them with a numerical simulation. You can use the jupyter notebook template on the webpage that guides you through the exercise (but you may also write it completely on your own, if you prefer).

a) *Single trajectory*

To begin with, we want to observe the behaviour of single realizations. For that, use the Gillespie algorithm to simulate the Birth-Death process. Note that you now have two possible events that can happen (whose rates depend on the current state of the system).

- Plot some trajectories until extinction for different values of the initial population size n_0, the birth rate λ and the death rate μ (for now choose $\mu > \lambda$). What do you observe?
- Now choose $\lambda > \mu$. Start with a λ that is only slightly larger than μ (e.g. $\lambda = 1.01$, $\mu = 1.0$). What do you observe? (If not done so, limit the time you simulate). Now increase λ a bit (e.g. to $\lambda = 1.05$) and try different n_0. What happens with the population size? What starts to happen to the time duration your simulation needs to finish? What could cause this drastic change in the performance of the Gillespie algorithm?

b) *Ensemble of trajectories*

Extend your program such that you are able to create an ensemble of trajectories (using the Gillespie algorithm).

- Simulate such an ensemble of different runs and from that calculate the mean of N, the number of non-degraded individuals, as a function of time. Note that the steps in different runs happen at different points in time. This complicates taking the mean at a certain point in time. In the provided notebook you find some tips how you may do that.
- Compare your numerical results for $\langle N(t) \rangle$ with the analytic calculations (plot both into one figure). If necessary, explore different values for the ensemble size until you get reasonable agreement between analytics and numerics. Choose $\mu > \lambda$. (You could for instance use $N(0) = 50$, $\lambda = 1.0$, $\mu = 1.1$.)
- Using your numerical results, obtain the extinction probability $p_0(t)$ as a function of time and compare it to the analytic formula obtained in the previous exercise (plot both into one figure). Do this for a case where $\mu > \lambda$ and a case where $\lambda > \mu$. (You could for instance use $N(0) = 50$, $\lambda = 1.0$, $\mu = 1.1$ and $N(0) = 10$, $\lambda = 1.01$, $\mu = 1.$)
- Using your numerical results, obtain the probability distribution for the extinction time. Do this for different parameters and compare it to what you sketched in the previous exercise.

Your solutions should be handed in by uploading them to Moodle by **Wednesday, 14^{th} May 2025, 10:00 am**.