

Can we analytically predict cell population properties based on single-cell properties?

Author: Yair Hein

Supervisors: Farshid Jafarpour, Ivan Kryven and René van Roij

Abstract

Many bacterial cells like *Escherichia coli* grow exponentially in size and divide symmetrically. The division time and exponential growth rate are not constant. Fluctuations in these single-cell properties affect the properties of populations of cells, such as the growth rate of the number of cells and the distribution of cell ages and sizes within a population. This paper is about finding the relations between single-cell behaviour and cell population properties. There is an existing single-cell behaviour model that comes short when describing the right correlations between growth rates of mother cells and daughter cells. We devise a new model that models bacterial growth rate in a way that is much more faithful to experimental data. We analytically derive population properties for this new model as well as for the existing model. This includes an analytical cell size distribution, something which up until this point has only been derived for a very simplistic single-cell division model.

Contents

1	Introduction	2
2	Background	5
2.1	Powell's relationship	5
2.2	Age distribution	7
2.3	Size distribution	8
2.4	Cell size control	9
3	Cell size control model	11
4	Discrete-step growth rate variability model results	12
4.1	General results	12
4.2	Normal distributions	13
4.3	Asymptotic behaviour	14
5	Ornstein-Uhlenbeck growth rate variability model	16
6	Ornstein-Uhlenbeck growth rate variability model results	17
7	Conclusions	19
8	Future work	19
A	Comparison of expected population and age distribution to other literature	20
B	Discrete-step growth rate variability population size and age density	22
B.1	General time-dependent formulas	22
B.2	Assuming normal distributions	23
B.3	Asymptotic total population	25
B.4	Asymptotic age distribution	26

C Ornstein-Uhlenbeck growth rate variability model	27
C.1 Time-dependent formulas	27
C.2 Asymptotic total population	30
C.3 Asymptotic size distribution	31
D Comparison of discrete-step growth rate and Ornstein-Uhlenbeck growth rate	32

1 Introduction

Qualitatively, reproduction in many single-celled organisms can be described as a fairly straight-forward process. A newborn cell grows until it reaches a size roughly twice its birth size. Then it splits in two and the process repeats. Understanding the actual quantitative properties regarding this simple process, such as cell size growth and division timing is a lot more complicated. Figure 1 is a classic image of the bacterium *Escheri-*



Figure 1: Electron micro-graphs of *E. coli* under differing growth conditions. [TW80]

cia coli, arguably one of the most well-studied and modelled organisms in biology [TW80]. We see different groups of cells, some fatter, others skinnier that all have the same genetic information. These differences were caused by the fact that they were grown under differing growth conditions. By growth conditions we mean factors such as the nutrient concentrations and temperature of their environment. Although the quantitative relations between growth conditions and single-cell behaviour remain largely a mystery, quantitative single-cell behaviour is surprisingly reproducible under constant growth conditions. An early example comes from the 1950s when Ole Maaløe and a group of biologists discovered that on average the size of *Salmonella* grows proportionally to its own size with a proportionality constant that only depends on the growth conditions [SMK58]. Thus in any medium at a constant temperature and with a homogeneous chemical composition, all cells in a population of cells belonging to the same species will on average grow exponentially with the same rate regardless of the ages and sizes of these cells. This relation is often referred to as the 'nutrient growth law' and it lies at the heart of many modern theories concerning single-cell and population growth. Note that this growth law is a phenomenon that emerges from more complicated internal biochemical cell processes involving protein concentrations. After a sudden change in growth conditions, some time is needed for the protein concentrations to reach steady state again. The nutrient growth law applies only during balanced growth which is when the protein concentrations are in a steady state [KMS58]. We are however only going to be concerned with models and experiments under constant growth conditions.

It is still important to keep in mind that the nutrient growth law is an emergent law emerging from more complex internal cell processes, much in the same vein as most classical laws in physics emerge from more complex quantum processes. In addition to the growth law, we will need some model that tells us when cells divide.

The 'timer model' first introduced by Powell in 1956 [Pow56] assumes that on average cells divide after some fixed time, which is the same for all cells independent of its size, or parents' division times. The actual division times will be independently distributed around this mean value. Since cells divide in two at division, this fixed average division time needs to correspond to the time it takes for a cell to double in size according to the growth law. An alternative characterization of the timer model is thus the assumption that cells on average divide after they have doubled in size with respect to the birth size. The 'sizer model' is another way to model division timings. In this model it is assumed that cells on average divide once they have reached a fixed size. In equilibrium the difference between these models lies in the way division time noise correlates between mother and daughter cells.

We will however assume that cell division timing follows the 'adder model' which tells us that cells will on average initiate division once they have gained some constant size with respect to their birth size [Ami14]. We will explain this model as well as the timer model in further detail in the background section.

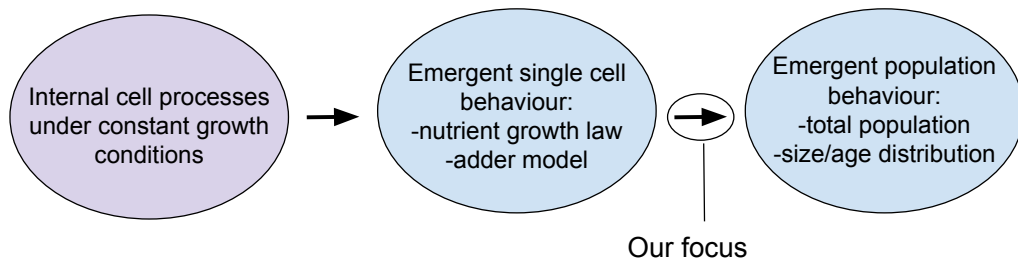


Figure 2: An overview of the relation between emergent phenomena of cell behaviour.

Once the single cell behaviour is known, one could run a simulation of a population of cells and observe several properties of this population. Such properties include the total number of cells, the distribution of ages and the distribution of sizes, each as a function of time. Size distributions and the total number of cells in a population can be directly measured in cell population experiments, so comparing said properties of simulated populations to actual experiments is an alternative way to validate a single cell model. As we shall later explain in further detail, single cell behaviour is random, so each simulation results in different population properties. In all of our models we will consider a cell population starting from one newborn cell. By averaging these population properties over a large number of simulations, one can however find some underlying 'expected' population properties. For any simulation, the measured population properties converge to their expected values with time.

The main focus of this thesis is to analytically derive the expected population properties, such as the total population, size distribution and age distribution as functions of time based on the adder model and nutrient growth law assumptions. These assumptions and their associated parameters such as the growth rate and division time variability emerge from internal processes and growth conditions that are outside of the scope of this thesis. An overview of the relationships we are interested in is given in figure 2. We will also find the asymptotic population properties, meaning the expected total population and age or size distribution when the simulation time is sent to infinity. These asymptotic expressions are much simpler than their dynamic time-dependent counterparts and more useful since in cell populations in many experimental setups, the population properties are approximately in their asymptotic state.

Up until this point we have been careful to only state the growth law and adder model in terms of 'average' cell behaviour because as mentioned before, single-cell behaviour is random. Division times and growth rate are best described by stochastic modelling. The exact sources of these forms of stochasticity lie within the internal cell processes. How these quantitatively affect the variability in division times and growth rate is still largely unknown, so we will mostly try to model any noise processes to match the variations observed in single-cell data [Jun+18]. Any choice made in the creation of these models will be a trade-off between generality, analytical solvability and faithfulness to experimental observations. In the case of this thesis for example, most analytical solutions only reduce to a simple form if we assume that sources of noise are small and/or normally distributed. Furthermore, in experiments, correlations show up between division times and growth rates of mother and daughter cells [Tah+15; Wan+10]. There are two sources of noise that contribute to the random nature of division times. One source comes from the division process and one from the randomness in growth rate. Both sources of noise are correlated between mother and daughter cells. The direct division time noise correlations are accounted for by using the adder model. That is, the noise in the difference in birth size and the size at division that is directly due to division process noise seems to be uncorrelated between mother and daughter cells. The correlation between growth rates of mother and daughter cells is of a different type. In principle the growth rate is a process that fluctuates around some equilibrium value in time. Newborn cells inherit the the growth rate of their mother cell at the time of division, which causes some correlations between the effective growth rates of mother and daughter cells. We will end up treating two slightly different models, one model where growth rate is constant for each cell and uncorrelated to the growth rate of other cells and one model where growth rate is modelled as a continuous stochastic process in time that also properly captures the right growth rate correlations observed in experiments [Wan+10]. Although continuous stochastic modelling is common practice in biology [PJI17], to the best of our knowledge a model with continuous stochastic growth rate and cell size regulation together has not been modelled before.

There are several benefits to finding analytical solutions for the total population or size distributions in terms of single cell behaviour. Besides helping to predict and understand population behaviour, it may also help infer single-cell quantities based on population quantities. Generally, large scale population quantities are easier to measure than single-cell quantities. Some of the distribution specific parameters especially, such as the growth rate variability may be hard to measure directly [Jun+18]. On the other hand, Ref. [LA17] shows that the difference between the average growth rate of a single cell and the population growth rate is a relatively simple function of this growth rate variability. This hints at the potential to deduce the single-cell growth rate variability from large scale measurements. Another example comes from Ref. [Jaf19] where the author found a relation between growth rate variability and oscillating patterns in the population growth rate in time. The growth rate inferred from some test data using this relation was in perfect agreement with direct measurements of the growth rate variability [Wan+10]. In our case we end up finding a simple formula for the size distribution of a population that depends solely on the cell size gain variability, by which we mean the variability in the difference between the size at division and the size at birth of some cell. Theoretically, this formula could be used to determine the cell size growth variability based on a single snapshot of the population. Directly measuring cell size growth variability would have to involve tracking numerous individual cells in time.

So far, analytic expressions of the age distribution have been found for the timer model, which assumes that division times of mother and daughter cells are independent [Jaf+18]. In this thesis we will for the first time derive analytical expressions of the population age distribution for the adder model instead of the timer model. Furthermore, there are no known analytical solutions for the size distribution yet. This thesis presents solutions for the size distribution of an adder model with a continuously modelled growth rate. We will also provide an expression for the asymptotic population growth rate for an adder model with a continuous growth rate, something that has not been done before either. Modelling the growth rate as a continuous process very closely resembles actual measurements of growth rate in time [Wan+10]. Furthermore, for each solution of age distributions, size distributions and total populations we will provide both dynamical time-dependent solutions and reduced solutions in the asymptotic large time limit.

In order to draw the link between single cell properties and emergent population properties, one first needs

to have an idea of what these population properties look like. In the background section we will derive and explain the known asymptotic population age distribution and size distribution behaviour for the simplest single cell model, namely the one where cells grow exponentially at a constant rate and divide after a fixed time or equivalently after a fixed size gain. We will also derive and explain the asymptotic population total and mention the asymptotic age distribution in the case of the timer model, where cells grow exponentially at a constant rate but divide after some randomly distributed time uncorrelated to other division times in the population.

We will then explain the shortcomings of the timer model. It is not only at odds with experimental data but it also leads to an interesting paradox when considering cell size in such a model, which we will explain in that section.

In the Cell size control model section we will be describing the single-cell model we will be using in this thesis, namely the adder model formulated as in Ref. [Ami14].

In the next section, we will first provide general expressions for the expected population total and age distribution in terms of unspecified division time distributions and briefly explain them. Then we can make these expressions more concrete by using division time distributions derived for the adder model and show how these expressions compare to simulated population totals and age distributions. We also provide the asymptotic population total and age distribution derived from the general time dependent expressions and briefly discuss them.

The adder model from Ref. [Ami14] does not properly model growth rate the way it behaves in experiments, which show that growth rate is a continuous process with non-negligible auto-correlations. We will introduce a new model based on the adder model from before, where we treat the growth rate like a continuous process instead of a discrete set of independent random variables. For this model we derive time-dependent expressions for the population total and size distribution and use these to find the asymptotic population total and size distribution. We then discuss these asymptotic results, compare them to simulations and results from the previous model.

In the final section we discuss our findings and mention how they could be used in further research.

2 Background

In experiments, cell populations commonly appear in time-independent equilibrium states. It is therefore of great importance to know the properties of populations in equilibrium. The size distribution, age distribution and population growth rate all converge to stationary distributions or constants one would encounter in such equilibrium states. In this section we will be deriving these stationary distributions for simplified cell division models and explain the results intuitively. After that we will discuss some known cell division models and explain their shortcomings while keeping the number of mathematical derivations to a minimum.

2.1 Powell's relationship

The effect of the single cell division times on the growth of the total population has first been investigated by Powell in 1956 [Pow56]. We will briefly explain his result and the intuition behind it, as it is a good introduction to modelling cell population dynamics and it is also used in the derivation of some of the analytical results found later in this thesis. Suppose cells divide independently of one another and that for any newborn cell, its age at division will be distributed randomly with probability density $g_\tau(\tau)$. Powell's relationship states that large cell populations grow exponentially as $N(t) \propto e^{k_\infty t}$ where k_∞ satisfies Powell's relationship, which reads

$$\int_0^\infty g_\tau(\tau) e^{-k_\infty \tau} d\tau = \frac{1}{2}. \quad (2.1)$$

Here g_τ is the density distribution function of single cell division times and k_∞ is the asymptotic exponential growth rate of the total cell population, meaning that at large times the total cell population follows $N(t) \propto$

$e^{k_\infty t}$. For example, if all cells divide at some fixed age $\bar{\tau}$, the division time probability density will be given by $g_\tau(\tau) = \delta(\tau - \bar{\tau})$. Filling this in into Powell's relationship results in the relationship

$$e^{-k_\infty \bar{\tau}} = \frac{1}{2}, \quad (2.2)$$

from which follows the expected result that $k_\infty = \frac{\ln(2)}{\bar{\tau}}$, hence the population exactly doubles in the time it takes a cell to divide. We shall later see that for any sharply peaked division time distributions, k_∞ will still be close to this value.

It is important to note that g_τ is not the division time distribution of a random cell taken from an existing population, but rather a distribution of division times along any lineage of offspring stemming from one isolated cell.

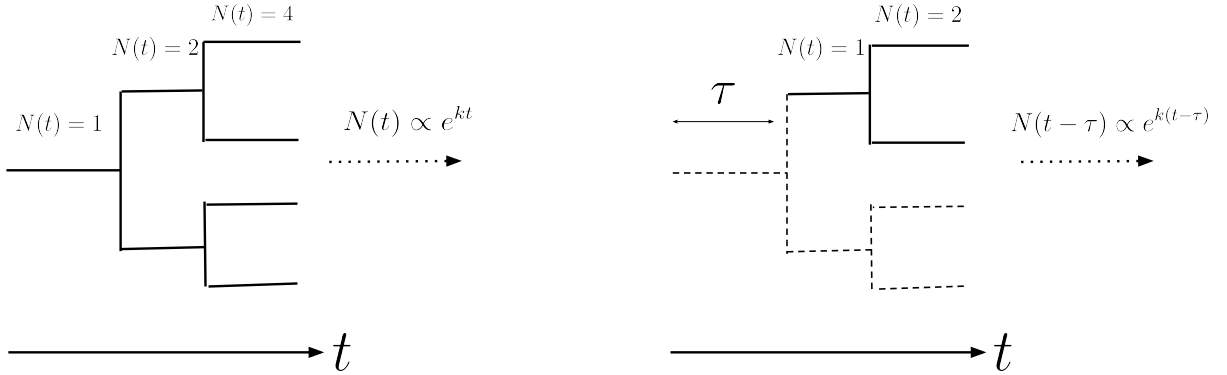


Figure 3: The left illustration sketches the situation where the first cell is born at time $t = 0$ and the right illustrates the situation where the first cell is born at time $t = \tau$.

We will now provide the intuition behind Powell's relationship in equation 2.1. Suppose each cell splits in two newborn cells once it has been alive for some division time τ drawn from the distribution g_τ . Since the rate of total divisions in the population is proportional to the population itself, it makes sense that at least asymptotically, the total population would grow exponentially with some unknown growth rate k_∞ . Suppose the first cell is born at time $t = 0$ if we now condition the first division time τ_1 to some τ , we get two new lineages whose first cell is born at τ as illustrated in figure 3. Hence

$$N(t|\tau_1 = \tau) = 2N(t - \tau) \quad (2.3)$$

If we now integrate the division time of the first cell τ_1 over all possible times using its density distribution function g_τ we get

$$N(t) = \int_0^\infty g_\tau(\tau) N(t|\tau_1 = \tau) d\tau = \int_0^\infty g_\tau(\tau) 2N(t - \tau) d\tau \quad (2.4)$$

In other words

$$\int_0^\infty g_\tau(\tau) \frac{N(t - \tau)}{N(t)} d\tau = \frac{1}{2} \quad (2.5)$$

Under the assumption that $N(t) \propto e^{k_\infty t}$ for large t , we obtain that $\lim_{t \rightarrow \infty} \frac{N(t - \tau)}{N(t)} = e^{-k_\infty \tau}$. If we take the limit of t going to infinity on both sides of equation 2.5, we obtain Powell's relationship as in equation 2.1. During this derivation we did however make one crucial assumption which does not hold in practice as argued before, namely the assumption that the division time of some cell is independent of the division time of its parents.

2.2 Age distribution

Let us explain how to derive the asymptotic population properties, namely the population age distributions and size distributions in some simplified models. By showing that these results correspond to special cases of the models we derive later in this thesis, we can hopefully make them more intuitive.

Assume a model where each cell grows exponentially at a constant rate \bar{k} and always divides at some fixed age $\bar{\tau}$. One could see this as the special case where the division time distribution satisfies $g_{\tau}(\tau) = \delta(\tau - \bar{\tau})$. In this model all birth sizes and division sizes will be constant as well. Because cells split in two at division, the size at division must always be twice the birth size. Suppose the birth size is Δ . Since at division, the cell has grown for a time $\bar{\tau}$ at rate \bar{k} we get that the size at division equals $e^{\bar{k}\bar{\tau}}\Delta$, which leads to the equation

$$2\Delta = e^{\bar{k}\bar{\tau}}\Delta \quad (2.6)$$

from which follows that \bar{k} and $\bar{\tau}$ must satisfy $\bar{\tau} = \frac{\ln(2)}{\bar{k}}$. Although this does not hold in the case where the division times are not constant, it is important to keep in mind that the division times are always roughly equal to the time it takes a cell to double in size. One may have noticed that in the previous section we derived from Powell's relationship 2.1 that for this model the asymptotic population growth rate k_{∞} satisfies the same relation and thus $k_{\infty} = \bar{k}$. This is only true in the constant division time model. In any other model where the growth rate is randomly distribution and \bar{k} is merely its average, there will be a difference between k_{∞} and \bar{k} .

Let us now show a simple derivation of the stationary age distribution for this problem. The derivations presented here are a simple case of the derivations from Ref. [Pow56] Suppose $G(\tau)$ is the stationary age distribution such that $G(\tau)d\tau$ is the fraction of cells in a population aged between τ and $\tau + d\tau$. We have that $N(t)G(\tau)d\tau$ now corresponds to the total number of cells at time t aged between τ and $\tau + d\tau$. Suppose $0 \leq \tau < \bar{\tau}$. Now look at the population a little later at time $t + dt$. Each cell from the set of cell previously aged between τ and $\tau + d\tau$ will now have aged by a time of dt as well. Since no cells in this set have divided or have been born in the meantime, the number of cells aged between τ and $\tau + d\tau$ at time t must be equal to the number of cells aged $\tau + dt$ to $\tau + dt + d\tau$ at time $t + dt$. This leads to the following equation

$$N(t + dt)G(\tau + dt)d\tau = N(t)G(\tau)d\tau \quad (2.7)$$

By dividing out $d\tau$ and making dt small this turns into the following differential equation

$$\partial_t N(t)G(\tau) + N(t)\partial_{\tau}G(\tau) = 0 \quad (2.8)$$

We are considering the stationary distribution, so we can assume t to be large, which means that we can use the relation that $N(t) \propto e^{k_{\infty}t}$ derived in the previous section. It follows that $\partial_t N(t) = k_{\infty}N(t)$ and thus we can rewrite equation 2.8 to

$$k_{\infty}G(\tau) + \partial_{\tau}G(\tau) = 0 \quad (2.9)$$

Which is the classic differential equation that solves for

$$G(\tau) \propto e^{-k_{\infty}\tau} \quad (2.10)$$

for all $0 \leq \tau < \bar{\tau}$. Because all cells in this model divide at $\bar{\tau}$ and are born at age 0, the age distribution must vanish outside of $0 \leq \tau < \bar{\tau}$. By now choosing the proper normalization, we find that

$$G(\tau) = \begin{cases} 2k_{\infty}e^{-k_{\infty}\tau} & 0 \leq \tau < \bar{\tau} \\ 0 & \text{otherwise} \end{cases} \quad (2.11)$$

A plot of this equation is given in figure 4.

Although this is a very idealized version of actual division behaviour, this result does capture the effect population growth has on age distribution. The fact that the population is ever increasing skews the population age distribution to favour younger cells over older cells, in a similar vein to how growing human populations tend to have more youth. Even though a model with perfectly constant division times would

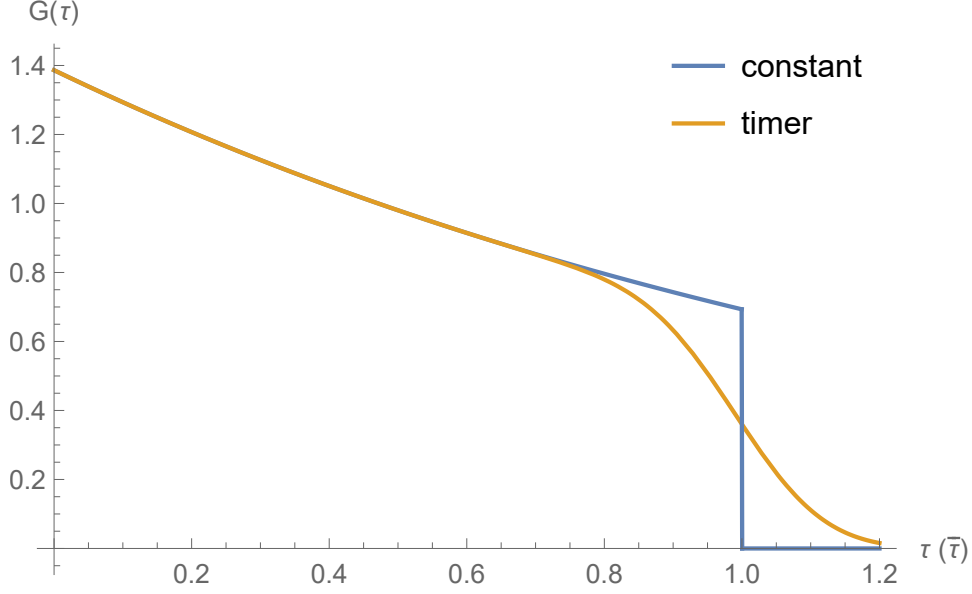


Figure 4: Plots of the constant division time stationary population age distribution as given in equation 2.11 and the timer model stationary population age distribution as in equation 2.12 for a normally distributed division time with standard deviation $\sigma_\tau = 0.1\bar{\tau}$

never converge to such a stationary age distribution, it's still worth calculating since one only needs infinitesimal growth rate variability to allow the age distribution to converge to its stationary distribution at some point in time [Jaf19]. This population growth skew factor of $2k_\infty e^{-k_\infty \tau}$ also shows up in age distributions derived for more complicated models. The population age distribution for the general timer model is given by [Pow56; Jaf+18; LA17]

$$G(\tau) = 2k_\infty e^{-k_\infty \tau} (1 - F(\tau)) \quad (2.12)$$

where $F(\tau) = \int_0^\tau g_\tau(\tau') d\tau'$ is the cumulative distribution function of the division time. The extra factor of $1 - F(\tau)$ is a decaying function corresponding to the decrease in old cells in the population due to division. The timer model stationary population age distribution in figure 4 can be thought of as follows: First until a cell age of roughly $\tau = 0.7\bar{\tau}$, divisions rarely happen and the decrease in age density is purely caused by the population growth. Then after $\tau = 0.7\bar{\tau}$ cells start dividing and the age density drops more sharply.

2.3 Size distribution

Let us see what the stationary size distribution is in this simple model. Like before, we choose a fixed birth size Δ . The size at division must now be 2Δ . Let $G(v)$ be the population size density distribution, meaning that $G(v)dv$ corresponds to the fraction of cells sized between v and $v + dv$. The total number of cells at time t sized between v and $v + dv$ is $N(t)G(v)dv$. Suppose $\Delta \leq v < 2\Delta$. Now consider this population a little later at time $t + dt$. Since their sizes have grown for a time of dt exponentially at a constant rate of $\bar{\kappa}$, we have that the previously mentioned set of cells is now the set of cells sized between $v e^{\bar{\kappa} dt}$ and $(v + dv) e^{\bar{\kappa} dt}$ at time $t + dt$, hence

$$N(t + dt)G(v e^{\bar{\kappa} dt}) e^{\bar{\kappa} dt} dv = N(t)G(v)dv \quad (2.13)$$

By dividing out dv and making dt small, this turns into a differential equation

$$\bar{\kappa} N(t)G(v) + \partial_t N(t)G(v) + v \bar{\kappa} N(t) \partial_v G(v) = 0 \quad (2.14)$$

The first term corresponds to the effect the direct cell growth has on the skew of the size distribution. Since cells grow faster as they get larger, they spend more time being small. The second term corresponds to the skew of the size distribution caused by population growth, analogous to the effect on age distributions. Like we argued before asymptotically we have that, $N(t) \propto e^{k_\infty t}$ where in this particular model it holds

that $k_\infty = \bar{\kappa}$, so $\partial_t N(t) = \bar{\kappa}N(t)$. If we fill this in into equation 2.14 we see that the two aforementioned effects, caused by direct cell growth and the population growth are exactly equal in magnitude. We obtain the following differential equation

$$2G(v) + v\partial_v G(v) = 0 \quad (2.15)$$

which has the family of solutions given by $G(v) \propto \frac{1}{v^2}$. If we now require that this distribution vanishes outside of the allowed range of sizes given by $\Delta \leq v < 2\Delta$ and normalize, we find that

$$G(v) = \begin{cases} \frac{2\Delta}{v^2} & \Delta \leq v < 2\Delta \\ 0 & \text{otherwise} \end{cases} \quad (2.16)$$

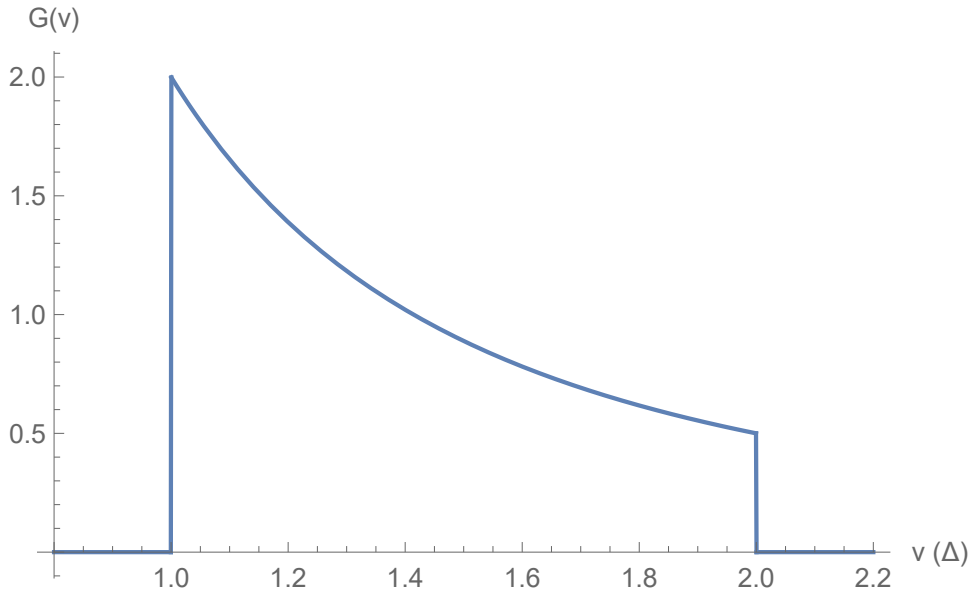


Figure 5: A plot of equation 2.16.

A plot of this equation is given in figure 5. In more realistic size distributions, when division sizes and birth size are random, the edges will be more smoothed out since some cells are born with a size smaller or larger than the equilibrium birth size.

An example of a size distribution of a cell population in equilibrium measured in an actual experiment is given in figure 6.

2.4 Cell size control

The model introduced by Powell is also called the timer model, since time since inception is the only factor that governs the division time. As mentioned in the introduction, single-cell observations however show that the cell volume of many bacteria grows exponentially with time [KMS58]. This is incompatible with the assumption that division times of cells and their parents are uncorrelated [Tah+15]. The idea is that if division times are uncorrelated, then so are the ratios between size of mother and daughter cells according to the growth law. If size ratios are uncorrelated and if there is any form of noise in these ratios, then this noise will accumulate over time, causing the size variability to diverge. Let us briefly explain this. Suppose that both independence of division times and the growth law are true, meaning we have some near-constant exponential growth rate κ and uncorrelated division times τ_n . Denote by v_{n-1} the birth size of some mother cell and v_n the birth size of its daughter cell. Note that v_n must be half the size of its mother cell at division. Since the mother cell has grown for a time of τ_n at the time of its division, it follows from the growth law that

$$v_n = \frac{1}{2}v_{n-1}e^{\kappa\tau_n} \quad (2.17)$$

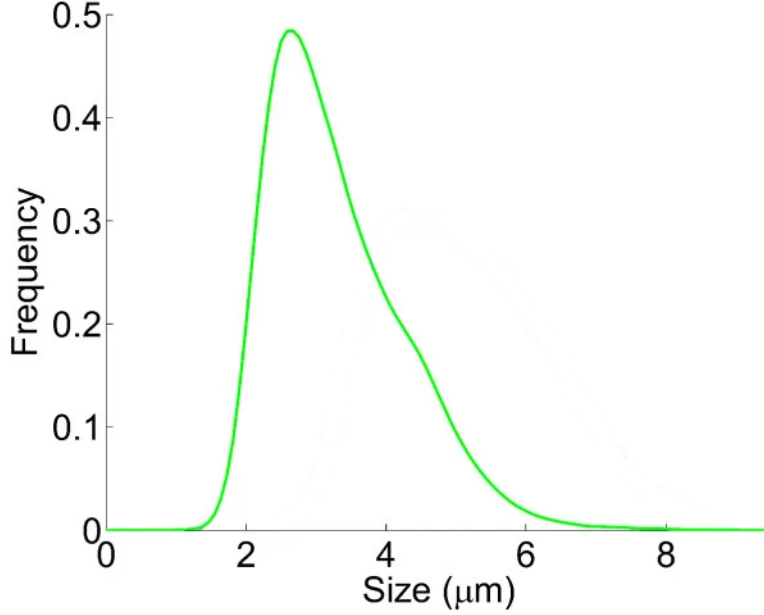


Figure 6: A cell size distribution of a population of *E coli* in equilibrium shown in Ref. [Rob+14] based on data from Ref.[Ste+05]. We were unable to obtain the data itself for data analysis.

and it recursively follows that the birth size of an n -th generation cell satisfies

$$v_n = \frac{1}{2^n} v_0 \exp \left[\kappa \sum_{i=1}^n \tau_i \right] \quad (2.18)$$

hence

$$\text{Var}(\ln(v_n)) = n\kappa^2 \sigma_\tau^2 \quad (2.19)$$

What is important here is that the right-hand side diverges as n goes to infinity. It follows that the width of any birth size distribution also becomes infinitely large, meaning that under the assumptions of this model any cell population would start producing any possible of size of cells, given enough generations. Experimental observations however tell us that all cells of the same type have sizes in the same order of magnitude, so the division times between cells and their parent cells must be correlated in some way.

One model that deals with this issue is the sizer model, which assumes that cells on average divide when they reach a fixed size. There is substantial evidence against the validity of this model as well [Cam+14]. Instead the model that seems to be most consistent with experimental data is the adder model [Tah+15], which dictates that cells grow exponentially and divide after they have gained a fixed amount of volume with respect to their size at birth. Mother cells born at a size larger than this fixed volume must now have daughter cells that are on average born smaller than the mother cells. Similarly, mother cells born at a size smaller than this fixed volume have daughter cells born larger than their mother cells. The cell birth sizes will now naturally tend to some equilibrium value. Any fluctuations in birth size dampen out more with each generation. In terms of the expressions used before to prove divergence of birth sizes in the adder model, we would get that

$$\text{Var}(\ln(v_n)) = \text{Var} \left(\kappa \sum_{i=1}^n \tau_i \right) \quad (2.20)$$

converges to some constant value as n goes to infinity, due to the correlations in the division times τ_i .

A formula for the age distribution in time already exists for the timer model [Jaf+18]. In this thesis, we will devise a similar formula that is instead based on the more realistic adder model. The framework we base our model on uses the model introduced in Ref. [Ami14] and further developed in Ref. [Jaf19].

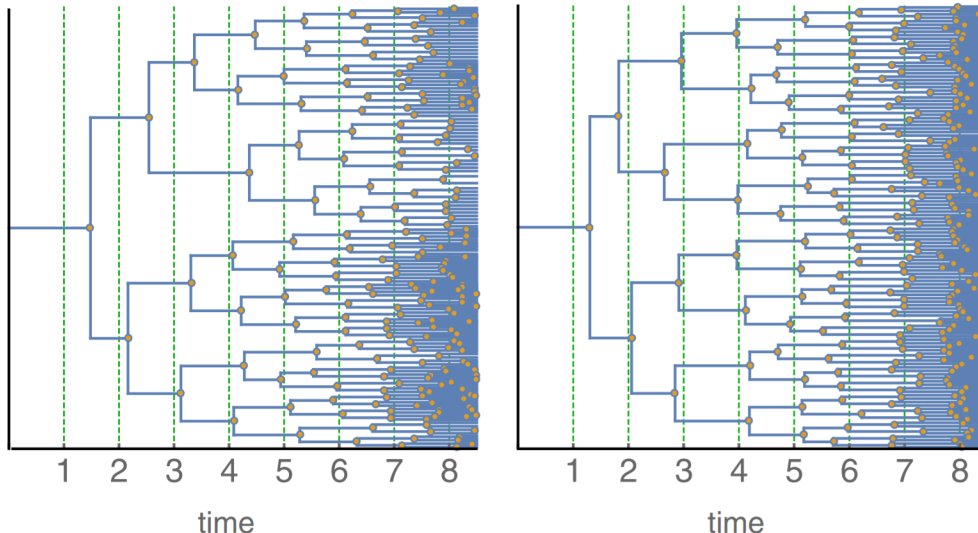


Figure 7: Both figures show lineage trees stemming from single cells born at $t = 0$ with constant growth rate. On the left we have a timer model where the division times per cell are uncorrelated. On the right we have a model with the same type of correlations as the adder model, which shows that the division times stay synchronized. This illustration was taken directly from Ref. [Jaf19] with the permission of the author.

We will go into the specifics of this framework in a later chapter. It was shown that in the case where all cells grow exponentially at the same rate, the variability of the time of a n -th generation cell division would not diverge [Jaf19]. To illustrate what this entails, we will consider the example of family trees. The age difference between you and your sibling is most likely small compared to a full lifetime. When you compare your age to the age of your first cousins, you're more likely to find larger differences. As we go further back, these differences start to diverge and we will get to a point where the generations don't line up anymore with lifetimes. In the adder model with constant growth rate however, the generations would always roughly line up no matter how many generations you would shift in time, see figure 7. In reality, there is slight variation in the growth rate. It was shown that the growth rate variability causes the division times of cells from one generation to desynchronize over time, albeit very slowly. This behaviour has been supported by experimental data [Jaf19]. Implementing growth rate variability is therefore necessary to get a model that is consistent with realistic long term behaviour of cell populations.

3 Cell size control model

In this section we will define and explain the discrete-step growth adder model. This version of the model was first introduced in Ref. [Ami14]. This model will partially account for correlations between division times of daughter and parent cells. For now we will assume that a n -th generation cell grows exponentially at a rate κ_n which is drawn independently from some distribution $\rho(\kappa)$. When a cell divides, its cell size halves. The birth size of some cell will therefore always be half the size of its parent at the time of division. Denote by v_n and the birth size of some cell and by v_{n-1} the birth size of its parent. Each cell attempts to divide when it reaches its target size, which depends on its birth size. This target size is given by

$$f(v_{n-1}) = 2\Delta^\alpha v_{n-1}^{1-\alpha} \quad (3.1)$$

Here α is some parameter between 0 and 1 and Δ is the equilibrium cell size. The reason we call Δ the equilibrium cell size is that it always holds that $f(\Delta) = 2\Delta$. The size of some cell at birth is half the size of its parent at division, so it will be born with size Δ again. When $\alpha = 1$, we see that $f(v_{n-1}) = 2\Delta$, hence the target size is some constant value. This is called the sizer model. When $\alpha = 0$ we see that $f(v_{n-1}) = 2v_{n-1}$.

Since growth is exponential, this corresponds to the situation where on average cells grow for some fixed time which is uncorrelated with cell sizes. This model is called the timer model. As mentioned before, both models fall short when describing actual cell populations. By choosing α in between 0 and 1 we can interpolate between the two models. Experimental data suggested that the adder model, where $f(v_{n-1}) = \Delta + v_{n-1}$ is the most accurate descriptor of actual cell division of *E. coli*. One finds that for $\alpha = 0.5$, equation 3.1 corresponds to the adder model up to second order in $v_{n-1} - \Delta$. In practice all birth sizes v_{n-1} are often sharply peaked around the equilibrium value Δ , so this becomes a good approximation. All of the simulations and numerical calculations will therefore use $\alpha = 0.5$. The reason we are using this model over the adder model is that equation 3.1 allows for better analytical solutions as we will later see. For some bacteria other than *E. coli*, their division behaviour may be best described by some other interpolation between the timer and sizer, with value $\alpha \neq 0.5$ [FCH17]. Although the only focus of this thesis is *E. coli*, keeping α unspecified could make it easier to generalize our results to other bacteria.

On average, a cell will attempt to divide again when reaching the size given by the target size from 3.1. This means that given some birth size v_{n-1} and growth rate κ_n , the average division time $\mathbb{E}[\tau_n|v_{n-1}, \kappa_n]$ must satisfy

$$f(v_{n-1}) = v_{n-1} e^{\kappa_n \mathbb{E}[\tau_n|v_{n-1}, \kappa_n]} \quad (3.2)$$

which can be rewritten as

$$\mathbb{E}[\tau_n|v_{n-1}, \kappa_n] = \frac{\ln(2)}{\kappa_n} + \alpha \frac{1}{\kappa_n} \ln \left(\frac{\Delta}{v_{n-1}} \right) \quad (3.3)$$

The next assumption is that the actual division times are distributed around this value where the noise random variables represented by ξ_n are centered around 0 and are independent of growth rates and noise variables from other generations.

$$\tau_n = \frac{\ln(2)}{\kappa_n} + \alpha \frac{1}{\kappa_n} \ln \left(\frac{\Delta}{v_{n-1}} \right) + \xi_n \quad (3.4)$$

Since the size of a cell at birth is half the size of its parent at division, we also have the recursion

$$v_n = \frac{1}{2} v_{n-1} e^{\kappa_n \tau_n} \quad (3.5)$$

We will assume that the first cell starts at the equilibrium size, so $v_0 = \Delta$. To recap, all ξ_n are independently distributed around 0 and all κ_n are independently distributed according to $\rho(\kappa)$. All τ_n and v_n now follow recursively from equations 3.4 and 3.5.

4 Discrete-step growth rate variability model results

4.1 General results

In this section, we will show analytically derived expressions for the total population and age distribution for the adder model described in the previous section, explain the intuition behind them and then show how they can be approximated in the case where noise variables are small and normally distributed. We will also compare them to simulated expected total populations and age distributions, simulated directly using the recursive relations that make up the model described in the previous section.

In appendix B we derive that the expected total population starting from one newborn cell at time $t = 0$ is given by

$$\mathbb{E}N(t) = 1 + \sum_{n=1}^{\infty} 2^{n-1} \mathbb{P}(t_n \leq t) \quad (4.1)$$

Where t_n is the random timing of an n -th generation division given by

$$t_n = \sum_{i=1}^n \tau_i \quad (4.2)$$

An example of equation 4.1 for approximated division times in an adder model is given in figure 8. Intuitively, equation 4.1 is not very hard to grasp. Since exactly one new cell is added to the population for each division, one can also obtain the total population at time t by counting how many divisions have happened before time t and adding the total population at time $t = 0$, which is one. Each term in the infinite summation on the right-hand side is the expectation value of the number of n -th generation divisions having occurred at time t . Here 2^{n-1} corresponds to the total number of possible n -th generation divisions and $\mathbb{P}(t_n \leq t)$ is the expectation value of each of these divisions having occurred at time t . So far, no assumptions about the nature of the division times have been made, so this is a very general result. The expression is in agreement with the population growth equation given in Ref. [Jaf19]. In Ref. [Jaf+18], an expression of the total expected population has been found for the timer model by solving a set of integral partial differential equations. Recall that by the timer model we mean that the division times are assumed to be independent random variables. In appendix A we show that in the timer model case, our formula is in perfect agreement with a special case of the total population formula from Ref. [Jaf+18].

We are also interested in finding the age distribution. By age distribution, we mean some function $G(t, \tau)$ such that after a simulation time of t , the total number of cells in the population aged between τ and $\tau + d\tau$ is $G(t, \tau)d\tau$. In appendix B we show that the expected age distribution is given by

$$G(t, \tau) = \frac{1}{\mathbb{E}N(t)} \left[\delta(t - \tau)\mathbb{P}(t_1 \leq t) + \sum_{n=1}^{\infty} 2^n g_{t_n}(t - \tau)\mathbb{P}(\tau \leq \tau_{n+1} | t_n = t - \tau) \right] \quad (4.3)$$

where g_{t_n} is the density distribution function of t_n . Here each term in the summation $g_{t_n}(t - \tau)\mathbb{P}(\tau \leq \tau_{n+1} | t_n)d\tau$ is the probability any random n -th generation cell is present in the population at time t and is aged between τ and $\tau + d\tau$. Per generation, we have 2^n such cells, hence the summation in equation 4.3 counts how many cells are expected to be present in the population at time t aged from τ to $\tau + d\tau$ over all generations. Dividing by the total population now results in the age density.

In appendix A we also show that under the timer model assumption, this is in agreement with a special case of the dynamic age distribution function given in Ref. [Jaf+18] and we discuss some of the differences. The most important advantage our formula in equation 4.3 has over the one in that paper is that it easily generalizes to allow for correlated division times.

Before we can visualize or use these functions at all, we need to make some assumptions about the distributions of the noise.

4.2 Normal distributions

In the case where noise variables ξ_n are independently normally distributed with mean 0 and standard deviation σ_ξ and κ_n are independently normally distributed around $\bar{\kappa}$ with standard deviation σ_κ , we show in appendix B that there is an approximate solution

$$\mathbb{E}N(t) = 1 + \sum_{n=1}^{\infty} 2^{n-1} \Phi \left(\frac{t - n\bar{\tau}}{\sqrt{\sigma_{\Xi_n}^2 + n\sigma_q^2}} \right), \quad (4.4)$$

where Φ is the standard normal cumulative distribution function, $\bar{\tau} \approx \frac{\ln(2)}{\bar{\kappa}} \left(1 + \frac{\sigma_\kappa^2}{\bar{\kappa}^2} \right)$ and

$$\sigma_{\Xi_n}^2 = \frac{1 - (1 - \alpha)^{2n}}{\alpha(2 - \alpha)} \sigma_\xi^2 \quad (4.5)$$

as well as

$$\sigma_q^2 = \mathbb{E} \left(\frac{\ln(2)}{\kappa_n} - \bar{\tau} \right)^2 \approx \frac{\ln(2)}{\bar{\kappa}^2} \frac{\sigma_\kappa^2}{\bar{\kappa}^2} \quad (4.6)$$

A log-plot of equation 4.4 alongside with a simulation is given in figure 8 for parameters used in Ref. [Jaf19]. The formula seems to be a good match. The oscillating patterns correspond to bursts of synchronized

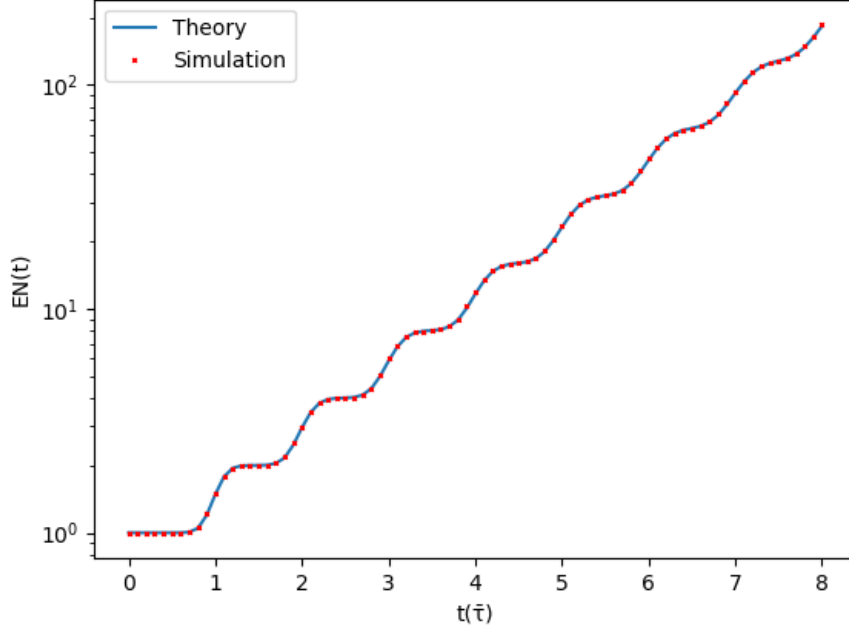


Figure 8: A log-plot of the expected total population as given in equation 4.4 against the average population over a large number of simulations with normally distributed κ_n and ξ_n where $\bar{\kappa} = \ln(2)$, $\sigma_\kappa = 0.07\bar{\kappa}$ and $\sigma_\xi = 0.1\bar{\tau}$

divisions caused by cells from the same generation dividing around the same time. As time passes, the divisions desynchronize, making the oscillating behaviour less and less prominent.

Furthermore the probability density of finding a cell with age τ at time t is given by

$$G(t, \tau) = \frac{1}{\mathbb{E}N(t)} \left[\delta(t - \tau) \Phi \left(\frac{\bar{\tau} - t}{\sqrt{\sigma_\xi^2 + \sigma_q^2}} \right) + \sum_{n=1}^{\infty} 2^n g_{t_n}(t - \tau) \Phi \left(\frac{-\frac{\alpha \sigma_{\Xi_n}^2}{\sigma_{\Xi_n}^2 + n \sigma_q^2} (t - n\bar{\tau} - \tau) + (\bar{\tau} - \tau)}{\sqrt{\sigma_\xi^2 + \sigma_q^2 + \frac{n \alpha^2 \sigma_q^2 \sigma_{\Xi_n}^2}{n \sigma_q^2 + \sigma_{\Xi_n}^2}}} \right) \right] \quad (4.7)$$

Note that in both this expression we have infinite summations. When you want to calculate the population or age distribution at time t , you can safely approximate them by taking all terms up to n in the order of $t/\bar{\tau}$, since all terms after that point go to zero very quickly.

Plots of equation 4.7 are given in figure 9 for simulation times ranging from $3\bar{\tau}$ to $4\bar{\tau}$. At such low simulation times, one can see the age waves caused by the fact that the division times and the ages of all cells are still strongly synchronized.

4.3 Asymptotic behaviour

By looking at figure 8 one might correctly guess that at large timescales, the logarithm of the expected population becomes more and more like a perfect straight line.

This suggests that the simulated expected population and equation 4.4 behave like exponential functions at large timescales, much like the timer model as demonstrated in the background section. This behaviour is further illustrated in figure 10 We analytically show in appendix B that for large t , equation 4.4 indeed behaves like an exponential function, namely

$$\mathbb{E}N(t) \propto e^{k_\infty t} \quad (4.8)$$

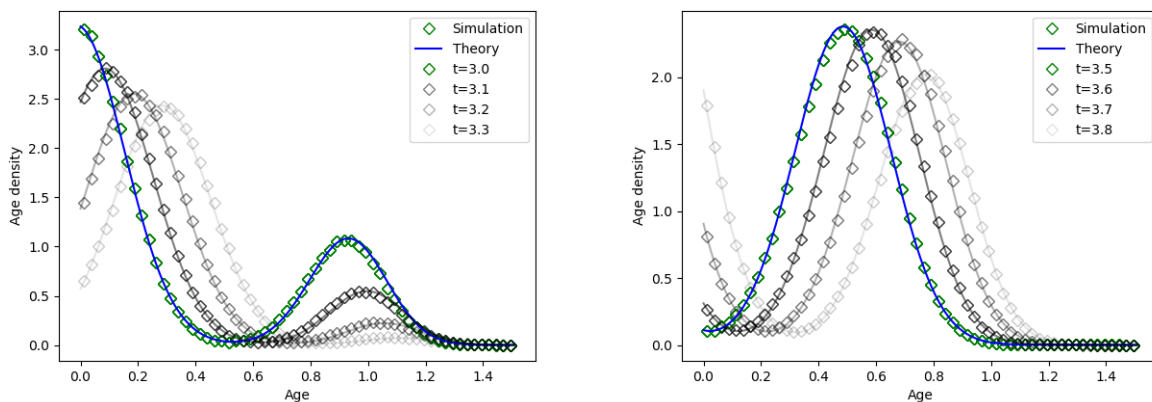


Figure 9: in these figures we see the simulated age density distributions plotted against equation 4.7 for different simulation times. The parameters used are $\bar{\kappa} = \ln(2)$, $\sigma_{\xi} = 0.1\bar{\tau}$ and $\sigma_{\kappa} = 0.07\bar{\kappa}$. All simulation time and age units are in terms of average division time $\bar{\tau}$.

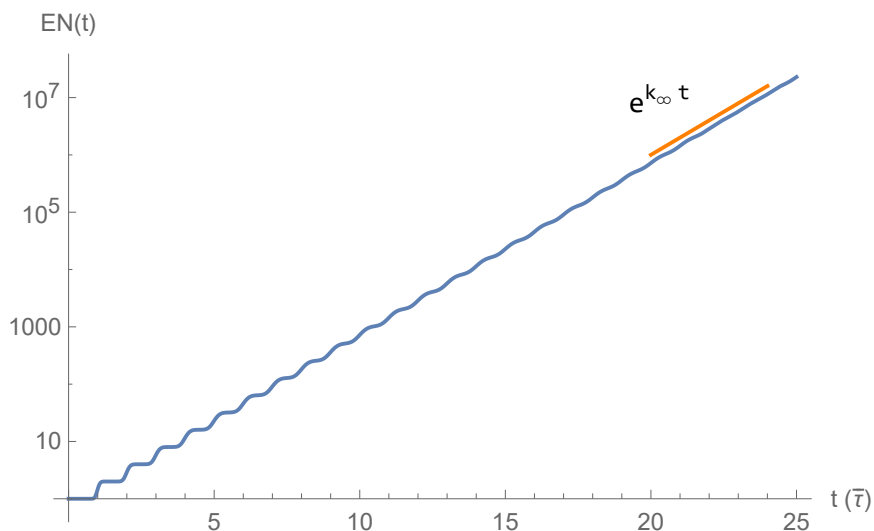


Figure 10: A log-plot of the analytically derived expected total population as given in equation 4.4 with $\bar{\kappa} = \ln(2)$, $\sigma_{\kappa} = 0.07\bar{\kappa}$ and $\sigma_{\xi} = 0.1\bar{\tau}$. A segment of the function $e^{k_{\infty}t}$ is shown for comparison where k_{∞} is given by equation 4.10

where the asymptotic growth rate k_{∞} is the constant satisfying the equation

$$\int 2^{-\frac{k_{\infty}}{\kappa}} \rho(\kappa) d\kappa = \frac{1}{2} \quad (4.9)$$

In the case where we assume ρ to be sharply peaked around $\bar{\kappa}$ with variance σ_{κ}^2 , this approximately solves for

$$k_{\infty} \approx \bar{\kappa} \left(1 - \left(1 - \frac{\ln(2)}{2} \right) \frac{\sigma_{\kappa}^2}{\bar{\kappa}^2} \right) \quad (4.10)$$

This result is in agreement with the asymptotic growth rates found in Ref. [Jaf19] and [LA20]. What's important here is that the difference between the asymptotic population growth rate and the average cell growth rate only depends on the variability of the cell growth rate and it is independent of other division time noise sources in the adder model. As for the age distribution, assuming that the random variables are

normally distributed we find that in the limit of large t

$$G(\tau) = 2k_\infty e^{-k_\infty \tau} \Phi \left(\frac{\bar{\tau} - \tau}{\sqrt{\sigma_q^2 + \frac{2}{2-\alpha} \sigma_\xi^2}} \right) \quad (4.11)$$

A good way to check the validity of this solution is to verify whether known solutions for limiting cases of the adder model coincide with special cases of equation 4.11. Recall that α is a parameter that interpolates between the timer model ($\alpha = 0$) and the sizer model ($\alpha = 1$). So by setting $\alpha = 0$ and ignoring the growth rate variability, thus setting $\sigma_q = 0$, we expect to recover the known stationary age distribution for the timer model where the division times are normally distributed around $\bar{\tau}$ with standard deviation σ_ξ . We obtain

$$G(\tau) = 2k_\infty e^{-k_\infty \tau} \Phi \left(\frac{\bar{\tau} - \tau}{\sigma_\xi} \right) \quad (4.12)$$

which is exactly in agreement with the expected formula for the timer model stationary age distribution as given in the background section in equation 2.12.

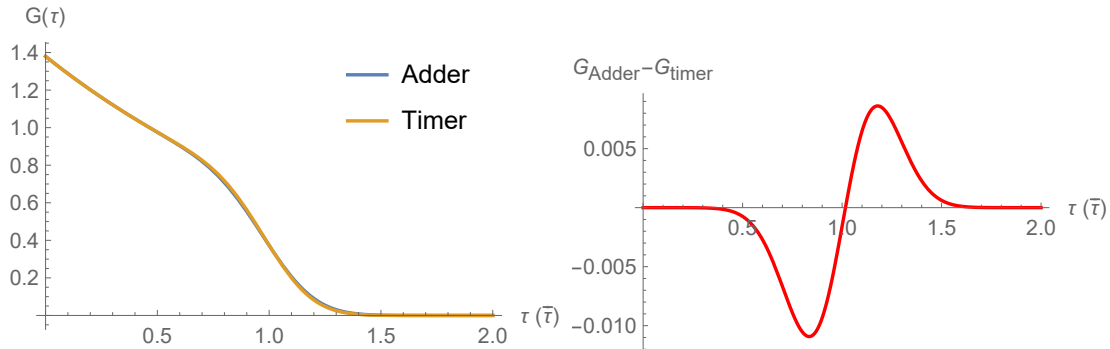


Figure 11: A comparison of the adder model asymptotic population age distribution from equation 4.11 without growth rate variability to the timer model asymptotic age distribution from equation 4.12, where $\bar{\kappa} = \ln(2)$ and $\sigma_\xi = 0.1\bar{\tau}$. The right plot shows the difference between the two distribution functions

A comparison between these two formulas is given in figure 11. In essence, equation 4.11 is not all that different from that of a timer model, only that range of ages cells divide at tends to be a little wider. The reason the cell division ages are spread out more, is because unlike in the timer model, division time noise from previous generations carries over to division time variability of new generations through birth size variability.

5 Ornstein-Uhlenbeck growth rate variability model

In all of the previous models, growth rates were drawn from some distribution with growth rates of mother and daughter cells being independent. Experimental data however suggests that growth rate correlation between mother and daughter cells are non-negligible [Wan+10]. These correlations can be understood as correlations caused by the fact that the growth rate process behaves like a continuous process, where daughter cells inherit the internal states from their mother cells at division. In this section we will model the growth rate as an Ornstein-Uhlenbeck process instead of assuming that each cell is assigned some random growth rate that remains constant during the life of that cell. For this model it will actually be easier to derive the size distribution but harder to derive the age distribution. The target size function f is still given by 3.1, but instead of adding time-additive noise, we will now add logarithmic size-additive noise. The division size of some cell born with size v_{n-1} will divide when it reaches the size

$$f(v_{n-1})e^{\eta_n} \quad (5.1)$$

where the size noise random variables η_n are centered around 0 and are independent of growth rate and noise variables from other generations. equation 5.1 results in the following recursive relationship between

cell birth sizes

$$v_n = \frac{1}{2} f(v_{n-1}) e^{\eta_n} = \Delta^\alpha v_{n-1}^{1-\alpha} e^{\eta_n} \quad (5.2)$$

In actual experiments, the growth rate changes continuously in time [Wan+10]. The important characteristics are that at each point in time, the growth rate can be assumed to be normally distributed and it continuously hovers around some mean value in time. The auto-correlation of this growth rate approximately decays exponentially with time. Now instead of assuming a growth rate that changes with each cell, we assume the growth rate λ_t to be governed by an Ornstein-Uhlenbeck process

$$d\lambda_t = \theta(\bar{\lambda} - \lambda_t)dt + \sigma_\lambda \sqrt{2\theta} dW_t \quad (5.3)$$

Here W_t is a Wiener process, $\bar{\lambda}$ is the time average growth rate, θ is the strength of the drift towards the average growth rate and σ_λ^2 can be shown to be the asymptotic variance of the solution λ_t . We assume λ_0 to be distributed according to the stationary distribution of the process, which is $\lambda_0 \sim N(\bar{\lambda}, \sigma_\lambda^2)$. One may think of θ as the decay rate of the auto-correlation function, since one can show that [Doo42]

$$\frac{Cov(\lambda_t, \lambda_{t+s})}{Var(\lambda_t)} = e^{-\theta s} \quad (5.4)$$

The discrete-step growth rate model discussed before is not necessarily incompatible with the assumption that growth rate changes continuously during the lifetime of a cell, since the cell growth rate from the previous model acts as an effective average of the growth rate process during the lifetime of that cell. The problem lies in the non-negligible correlations between the effective growth rates of the mother and daughter cells, caused by the auto-correlations of the underlying continuous growth rate process. In the case where we let the auto-correlations vanish by sending θ to infinity while keeping the cell-average effective growth rate variability constant, we expect to obtain a model equivalent to the discrete-step growth rate variability model. In appendix D we show that the asymptotic population growth rates are in agreement in this limit of vanishing auto-correlations.

6 Ornstein-Uhlenbeck growth rate variability model results

In appendix C we derive dynamic expressions for the total population and size distribution and then use those to find the asymptotic population properties. It turns out that for large t the expected population again becomes an exponential function

$$\mathbb{E}N(t) \propto e^{\lambda_\infty t} \quad (6.1)$$

where the asymptotic population growth rate λ_∞ is given by

$$\lambda_\infty = \bar{\lambda} + \frac{\sigma_\lambda^2}{\theta} \quad (6.2)$$

Just like with the discrete step growth rate model, the difference between the time average growth rate and the population growth rate is independent of direct division noise. In this case that is size additive division noise and in the other model that is the direct time additive division noise. Like we mentioned before, in the limit of vanishing auto-correlations in the Ornstein-Uhlenbeck process governing the growth rate ($\theta/\bar{\lambda} \gg 1$), we may expect the model to behave similarly to the discrete step growth rate with independent growth rates. In appendix D we show that in this limit, equation 6.2 is indeed in agreement with equation 4.10.

Furthermore we find that the asymptotic size distribution can be expressed as

$$G(v) = \frac{2\Delta}{v^2 \mathbb{E}e^{-\delta u_\infty}} \mathbb{P}(\ln(v/\Delta) - \ln(2) \leq \delta u_\infty \leq \ln(v/\Delta)) \quad (6.3)$$

Where δu_∞ is a random variable satisfying

$$\delta u_\infty \sim \sum_{i=1}^{\infty} (1-\alpha)^{i-1} \eta_i \quad (6.4)$$

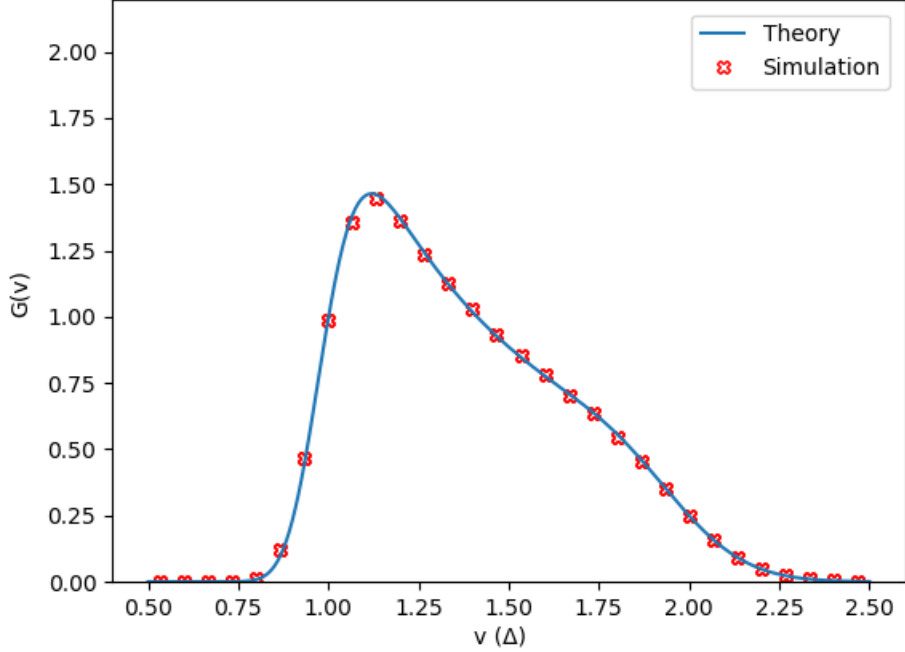


Figure 12: A comparison of the analytic asymptotic population size distribution of the Ornstein-Uhlenbeck growth rate adder model from equation 6.5 to a simulated size distribution for the discrete growth rate adder model with a simulation time of 2000 average division times. We used $\bar{\kappa} = \ln(2)$, $\sigma_{\kappa} = 0.07\bar{\kappa}$, $\sigma_{\eta} = \bar{\kappa}\sigma_{\xi} = 0.1\bar{\kappa}\bar{\tau}$

If we remove all division size noise by setting $\sigma_{\eta} = 0$ our model becomes equivalent to the constant division time model treated in the background section. In this scenario we get that $\delta u_{\infty} = 0$ with full probability and thus equation 6.3 turns into equation 2.16 as expected.

Under the assumption that η_n are normally distributed, we can simplify equation 6.3 to something a little more useful, namely

$$G(v) = \frac{2\Delta}{v^2} e^{-\frac{1}{2}\sigma_{u_{\infty}}^2} \left(\Phi\left(\frac{\ln(v/\Delta)}{\sigma_{u_{\infty}}}\right) - \Phi\left(\frac{\ln(v/\Delta) - \ln(2)}{\sigma_{u_{\infty}}}\right) \right) \quad (6.5)$$

where

$$\sigma_{u_{\infty}}^2 = \frac{1}{(2-\alpha)\alpha} \sigma_{\eta}^2 \quad (6.6)$$

Note that the asymptotic size distribution is completely independent of any parameters governing the growth rate process λ_t , namely the time-average growth rate $\bar{\lambda}$, the growth rate variability σ_{λ} and the growth rate auto-correlation decay rate θ . In figure 12 We show equation 6.5 against a simulation. We have not had the time to properly implement a simulation of the Ornstein-Uhlenbeck growth rate model so instead we have compared the asymptotic size distribution to the size distribution of the discrete growth rate model. The fact that figure 12 shows that these size distributions match so well, supports the universality and applicability of the Ornstein-Uhlenbeck growth rate asymptotic size distribution. It is worth noting that visually, this theoretical size distribution looks very similar to the experimental size distribution from Ref. [Rob+14] using data from [Ste+05] shown in figure 6.

We have treated two versions of the adder model, one where growth rate changes continuously and one where growth rate is constant per cell. For the continuous growth rate model, we have only managed to

derive the size distribution and for the discrete model, we only have the age distribution, so unfortunately we cannot compare those.

7 Conclusions

First we introduced the adder model from Ref. [Ami14], which we called the discrete-step growth rate adder model. Though its asymptotic population growth already been studied, the dynamic time-dependent expected population total and age distribution formula are new as well as the asymptotic age distribution. Asymptotic age distribution formulas have been studied [Has+16; LA17], But there is no asymptotic age distribution in terms of single-cell parameters yet for the adder model.

We then devised an new model where growth rate is treated as an Ornstein-Uhlenbeck process. This way of modelling growth rate is much more faithful to experimental data than other existing single-cell models. As far as we know, there are no other single-cell models that model growth rate as a continuous process. For this model we found time-dependent as well as asymptotic formulas for the population total and cell size distribution. The asymptotic growth rate we found always increases with growth rate variability if the time-average cell growth rate is kept constant.

The asymptotic population cell size distribution is interestingly enough independent of any parameters governing the cell growth rate process, meaning the average growth rate, growth rate variability and auto-correlation decay rate. This hints at the notion that the asymptotic size distribution is completely independent of the growth rate process. The size distribution only depends on average birth size and division size variability. There are no other known closed formulas for the asymptotic size distribution yet.

8 Future work

As mentioned before, the asymptotic age distribution is something that has been investigated on several occasions too, and not just for the timer model [Has+16; LA17]. No such theories give population age distributions directly in terms of single-cell parameters yet. Looking into how our asymptotic population age distribution compares to such theories may lead to new insights.

We have shown that the asymptotic population growth rate of the Ornstein-Uhlenbeck growth rate model is in agreement with the growth rate of the discrete step growth rate in the limit of vanishing auto-correlations. In other literature, a discrete-step growth rate model with growth rate correlations has been investigated [LA20]. It may be worth investigating to what extent their results are in agreement with the asymptotic growth rate for the Ornstein-Uhlenbeck growth rate model with non-vanishing auto-correlation.

Size distribution data exist [Yos+14; Rob+14; Ste+05], but we have not gained access and analysed it yet. Ideally one would fit our formula to a set of data for which single cell parameters are known too. If the parameters deduced from the fit match those of the direct measurements, it would not only support the validity of our model but it would also show that the formula can be used as an alternative way of estimating single-cell parameters based on measurements of population properties.

A Comparison of expected population and age distribution to other literature

In this section we will demonstrate that our general formulas for the expected population and age distribution in equations 4.1 and 4.3 respectively are in agreement with the results from Ref. [Jaf+18]. Here, general expressions for the total population and age distribution in time are derived for the timer model by modelling the population through differential equations. They also considered a more general case where stalker cells were taken into the model. We will ignore the existence of stalker cells and use the specific case of their model where cell divide symmetrically. Let us first restate their terminology and results and then restate them in terms of the terminology used in our thesis.

- $P(\tau) = g_{\tau_1}(\tau)$ is the probability density such that $P(\tau)d\tau$ is the probability of cell living to age τ , then dividing between ages τ and $\tau + d\tau$
- $f(\tau) = 1 - \mathbb{P}(\tau_1 \leq \tau)$ is the probability of a cell living at least up age τ
- $\alpha(\tau) = P(\tau)/f(\tau) = g_{\tau_1}(\tau)/(1 - \mathbb{P}(\tau_1 \leq \tau))$ is the division rate. This means that a cell aged τ as a probability of $\alpha(\tau)d\tau$ of dividing before reaching age $\tau + d\tau$
- $N(t) = \mathbb{E}N(t)$ is the expected population
- $n(t, \tau) = G(t, \tau)\mathbb{E}N(t)$ is the density of cells aged τ at time t , meaning that the number of cells at time t aged between τ and $\tau + d\tau$ is equal to $n(t, \tau)d\tau$
- ν is the number of offspring per cell. In Ref. [Jaf19] it is kept unspecified, but we have set it to $\nu = 2$
- $n^{(0)}(\tau) = n(0, \tau)$ is the age density of the starting population. In Ref. [Jaf19] it is unspecified, but we considered only the specific case where there is one cell aged $\tau = 0$ at time $t = 0$, hence we set $n^{(0)}(\tau) = \delta(\tau)$

The total population satisfies

$$N(t) = \mathcal{L}^{-1} \left\{ \tilde{f}(s) \frac{\tilde{n}_1(s)\nu}{1 - \tilde{P}(s)\nu} \right\} (t) + \int_t^\infty n^{(0)}(\tau - t) \frac{f(\tau)}{f(\tau - t)} d\tau \quad (\text{A.1})$$

with

$$n_1(t) = \int_t^\infty n^{(0)}(\tau - t) \frac{f(\tau)}{f(\tau - t)} \alpha(\tau) d\tau \quad (\text{A.2})$$

and where

$$\tilde{f}(s) := \int_0^\infty f(t) e^{-st} dt \quad (\text{A.3})$$

denotes the forward Laplace transform and \mathcal{L}^{-1} the inverse Laplace transform. In equation A.1 the right term corresponds to all the cells from the starting population that have not divided yet at time t and the left term corresponds to all the other cells at time t . Since we set $n^{(0)}(\tau) = \delta(\tau)$, the right term on the right-hand side of equation A.1 can be simplified to just $f(t)$, which is precisely the expected number of cells from a starting population of one cell at $t = 0$ to have remained undivided up to time t . Similarly we now also have that $n_1(t) = f(t)\alpha(t) = P(t)$, so

$$N(t) = \mathcal{L}^{-1} \left\{ \tilde{f}(s) \frac{\tilde{P}(s)\nu}{1 - \tilde{P}(s)\nu} \right\} (t) + f(t). \quad (\text{A.4})$$

By Laplace transforming both sides and rewriting the equation, one can show that this is precisely equivalent to

$$\nu \tilde{P}(s) \tilde{N}(s) = \tilde{N}(s) - \tilde{f}(s). \quad (\text{A.5})$$

Through an inverse Laplace operation on both sides, we get

$$\nu \int_0^t N(t-s) P(s) ds = N(t) - f(t). \quad (\text{A.6})$$

Rewriting this in terms of the expressions used in our thesis, we see that this is equivalent to

$$2 \int_0^t \mathbb{E}N(t-\tau)g_{\tau_1}(\tau)d\tau = \mathbb{E}N(t) - \mathbb{P}(\tau_1 \leq \tau) + 1. \quad (\text{A.7})$$

This is precisely an equation our formula for the total population in equation 4.1 has been shown to satisfy during the derivation of the asymptotic growth rate in equation B.35.

Analogously, one could verify that our general formula for the age distribution in equation 4.3 satisfies the general solution given in Ref. [Jaf19]. Alternatively we can show that our solution directly satisfies the differential equations devised in Jafarpour's paper. These equations are

$$\partial_t n(t, \tau) + \partial_\tau n(t, \tau) = -\alpha(\tau)n(t, \tau) \quad (\text{A.8})$$

and

$$n(t, 0) = \nu \int_0^\infty n(t, \tau)\alpha(\tau)d\tau \quad (\text{A.9})$$

So in terms of our thesis's terminology, we want the age density from equation 4.3 to satisfy

$$(\partial_t + \partial_\tau)G(t, \tau)\mathbb{E}N(t) = -\frac{g_{\tau_1}(\tau)}{1 - \mathbb{P}(\tau_1 \leq \tau)}G(t, \tau)\mathbb{E}N(t) \quad (\text{A.10})$$

and

$$G(t, 0)\mathbb{E}N(t) = 2 \int_0^\infty G(t, \tau)\mathbb{E}N(t)\frac{g_{\tau_1}(\tau)}{1 - \mathbb{P}(\tau_1 \leq \tau)}d\tau \quad (\text{A.11})$$

for $t > 0$. In the case of the timer model, meaning uncorrelated division times, our solution for the age density distribution in equation 4.3 reduces to

$$G(t, \tau) = \frac{1 - \mathbb{P}(\tau_1 \leq \tau)}{\mathbb{E}N(t)} \left[\delta(t - \tau) + \sum_{n=1}^\infty 2^n g_{t_n}(t - \tau) \right]. \quad (\text{A.12})$$

We will now try to see if this expression satisfies the aforementioned equations. Filling it in into the left-hand side of equation A.10 gives

$$(\partial_t + \partial_\tau)G(t, \tau)\mathbb{E}N(t) = (\partial_t + \partial_\tau)(1 - \mathbb{P}(\tau_1 \leq \tau)) \left[\delta(t - \tau) + \sum_{n=1}^\infty 2^n g_{t_n}(t - \tau) \right] \quad (\text{A.13})$$

$$= -g_{\tau_1}(\tau) \left[\delta(t - \tau) + \sum_{n=1}^\infty 2^n g_{t_n}(t - \tau) \right] \quad (\text{A.14})$$

$$= -\frac{g_{\tau_1}(\tau)}{1 - \mathbb{P}(\tau_1 \leq \tau)}G(t, \tau)\mathbb{E}N(t), \quad (\text{A.15})$$

hence the equation is satisfied. Let us now fill in equation A.12 into the right-hand side of equation A.11, giving that for all $t > 0$ we have

$$2 \int_0^\infty G(t, \tau)\mathbb{E}N(t)\frac{g_{\tau_1}(\tau)}{1 - \mathbb{P}(\tau_1 \leq \tau)}d\tau = 2 \int_0^\infty g_{\tau_1}(\tau) \left[\delta(t - \tau) + \sum_{n=1}^\infty 2^n g_{t_n}(t - \tau) \right] d\tau \quad (\text{A.16})$$

$$= 2g_{\tau_1}(t) + 2 \sum_{n=1}^\infty 2^n g_{t_{n+1}}(t) \quad (\text{A.17})$$

$$= \sum_{n=1}^\infty 2^n g_{t_1}(t) \quad (\text{A.18})$$

$$= (1 - \mathbb{P}(\tau_1 \leq t)) \left[\delta(t) + 2 \sum_{n=1}^\infty 2^n g_{t_{n+1}}(t) \right] \quad (\text{A.19})$$

$$= G(t, \tau)\mathbb{E}N(t), \quad (\text{A.20})$$

hence both equations are satisfied. The solutions from [Jaf19] are given for an arbitrary number of offspring ν , initial total population and age distribution $n(0, \tau)$. In our solution, the number of offspring can easily be generalized to an arbitrary number of offspring ν instead of just 2 without any extra calculations. It is not as straightforward to generalize our solution to handle arbitrary initial age distributions. The solution from Ref. [Jaf19] also has the benefit that the more elaborate formula includes the possibility of modelling stalker cells. When it comes to actually using the solution to numerically calculate age distributions or population sizes, both provide useful but alternative non-closed form expressions to approximate the solution. One involves an inverse Laplace transform, the other some infinite sum that quickly converges after a given term that depends on the time scale used. The big advantage our solution has however is that unlike the solution from Ref. [Jaf19], it allows for correlated division times, which is necessary when modelling the adder model instead of the timer model.

B Discrete-step growth rate variability population size and age density

B.1 General time-dependent formulas

If we are only interested in cell population numbers and cell ages, we can fully characterize an experiment or simulation by the timings of each division. So far we have only considered a single lineage of cells at a time. In practice a new lineage is created with every division. In order to distinguish different divisions that happen in the same generation, we are going to introduce some new notation. The first cell has two children, labelled 0 and 1. The former then has two children again which we can distinguish by labelling them $(0, 0)$ and $(0, 1)$. Each n -th generation division time or noise variable can be fully characterized by a binary vector $j = (j_1, \dots, j_{n-1})$ of length $n - 1$. For notational simplicity we will write

$$t_n^j = t_{n, (j_1, \dots, j_{n-1})}. \quad (\text{B.1})$$

If j is a vector longer than length $n - 1$ we will also accept the definition above, trimming j down to just its first $n - 1$ entries. In order to find the total number of cells alive at a given time t , we need to count the number of cells (j_1, \dots, j_n) whose birth time t_n^j happened before t and whose division time t_{n+1}^j happened after t . In other words

$$N(t) = \mathbb{1}\{t \leq t_1\} + \sum_{n=1}^{\infty} \sum_{j \in \{0,1\}^n} \mathbb{1}\{t_n^j \leq t < t_{n+1}^j\}. \quad (\text{B.2})$$

This can be algebraically rewritten to

$$N(t) = 1 + \sum_{n=1}^{\infty} \sum_{j \in \{0,1\}^{n-1}} \mathbb{1}\{t_n^j \leq 1\}. \quad (\text{B.3})$$

Since exactly one extra cell is created with every division that happens, counting the total number of divisions that happened before time t and adding the initial population should also give you the total population, so this formula makes sense.

A plot of this counting function for one simulation is given in figure 13.

Even though all division times within one generation are heavily correlated, they still have the same distribution. We can therefore abuse the linearity of expectations to obtain the expectation value of the total population

$$\mathbb{E}N(t) = 1 + \sum_{n=1}^{\infty} \sum_{j \in \{0,1\}^{n-1}} \mathbb{P}(t_n \leq t) = 1 + \sum_{n=1}^{\infty} 2^{n-1} \mathbb{P}(t_n \leq t), \quad (\text{B.4})$$

which is equation 4.1.

Next we define some function that counts the total number of cells that are younger than some age τ . The idea is that the average of this function divided by the average population gives us a cumulative distribution function of the population age. By taking the derivative we will then obtain a population age

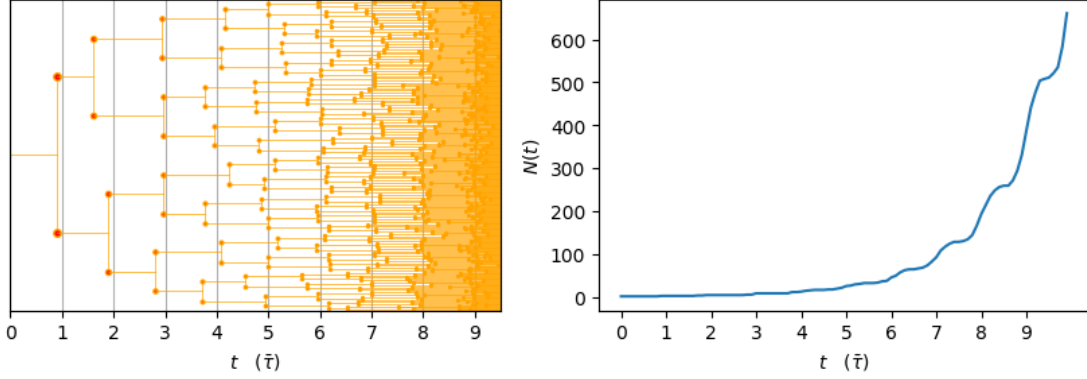


Figure 13: On the left side there is a visualisation of one simulation of an adder model with discrete growth rate variation. Each node corresponds to a division. On the right side is the corresponding total population of the same simulation in time.

density distribution function. A cell born at t_n^j has age $t - t_n^j$, so it is younger than τ if and only if $t - t_n^j \leq \tau$. The total number of cells younger than τ at time t is therefore given by the following expression.

$$T(t, \tau) = \mathbf{1}\{t \leq t_1, t \leq \tau\} + \sum_{n=1}^{\infty} \sum_{j \in \{0,1\}^n} \mathbf{1}\{t - \tau \leq t_n^j \leq t < t_{n+1}^j\}, \quad (\text{B.5})$$

hence

$$\mathbb{E}T(t, \tau) = \mathbf{1}\{t \leq \tau\} \mathbb{P}(t \leq t_1) + \sum_{n=1}^{\infty} \sum_{j \in \{0,1\}^n} \mathbb{P}(t - \tau \leq t_n \leq t < t_{n+1}) \quad (\text{B.6})$$

$$= \mathbf{1}\{t \leq \tau\} \mathbb{P}(t \leq t_1) + \sum_{n=1}^{\infty} 2^n \mathbb{P}(t - \tau \leq t_n \leq t < t_{n+1}). \quad (\text{B.7})$$

The probability density of finding a cell aged τ at time t is now given by

$$G(t, \tau) = \frac{d}{d\tau} \frac{\mathbb{E}T(t, \tau)}{\mathbb{E}N(t)} = \frac{1}{\mathbb{E}N(t)} \left[\delta(t - \tau) \mathbb{P}(t \leq t_1) + \sum_{n=1}^{\infty} 2^n \frac{d}{d\tau} \mathbb{P}(t - \tau \leq t_n \leq t < t_{n+1}) \right]. \quad (\text{B.8})$$

Now note that we can rewrite each of the terms in the infinite summations as follows

$$\frac{d}{d\tau} \mathbb{P}(t - \tau \leq t_n \leq t < t_{n+1}) d\tau \quad (\text{B.9})$$

$$= \mathbb{P}(t - \tau - d\tau \leq t_n < t - \tau, t < t_{n+1}) \quad (\text{B.10})$$

$$= \mathbb{P}(t < t_{n+1} | t - \tau - d\tau \leq t_n < t - \tau) \mathbb{P}(t - \tau - d\tau \leq t_n < t - \tau) \quad (\text{B.11})$$

$$= \mathbb{P}(t < t_{n+1} | t_n = t - \tau) g_{t_n}(t - \tau) d\tau \quad (\text{B.12})$$

$$= \mathbb{P}(\tau < \tau_1 | t_n = t - \tau) g_{t_n}(t - \tau) d\tau \quad (\text{B.13})$$

filling this in into equation B.8 results in equation 4.3.

B.2 Assuming normal distributions

In order to move any further with the population or age distribution formulas we derived before, we need to find the distributions of the timings of the divisions t_n . What we have are equations 3.4 and 3.5 which can be rewritten as simpler equations involving logarithmic size scales, namely

$$\kappa_n \tau_n = \ln(2) - \alpha \delta u_n + \kappa_n \xi_n \quad (\text{B.14})$$

and

$$\kappa_n \tau_n = \ln(2) + \delta u_n - \delta u_{n-1}, \quad (\text{B.15})$$

where $\delta u_n = u_n - \bar{u}$ with $u_n = \ln(v_n)$ and $\bar{u} = \ln(\Delta)$. One can combine equations B.14 and B.15 to obtain

$$\delta u_n = (1 - \alpha) \delta u_{n-1} + \kappa_n \xi_n \quad (\text{B.16})$$

and since we set $u_0 = \ln(v_0) = \bar{u}$, the above recursive formula can be solved to give

$$\delta u_n = \sum_{i=1}^n (1 - \alpha)^{n-i} \kappa_i \xi_i. \quad (\text{B.17})$$

Assuming that both ξ_n and $\delta \kappa_n$ are small, we can within reasonable approximation remove all cross terms between the different types of noise and we obtain that

$$t_n = \sum_{i=1}^n \tau_i \quad (\text{B.18})$$

$$= \sum_{i=1}^n \frac{\ln(2)}{\kappa_i} + \frac{\delta u_i - \delta u_{i-1}}{\kappa_i} \quad (\text{B.19})$$

$$\approx \sum_{i=1}^n \frac{\ln(2)}{\kappa_i} + \frac{\delta u_i - \delta u_{i-1}}{\bar{\kappa}} \quad (\text{B.20})$$

$$= \frac{\delta u_n}{\bar{\kappa}} + \sum_{i=1}^n \frac{\ln(2)}{\kappa_i} \quad (\text{B.21})$$

$$= \sum_{i=1}^n (1 - \alpha)^{n-i} \xi_i + \sum_{i=1}^n \frac{\ln(2)}{\kappa_i}. \quad (\text{B.22})$$

This can be split into a constant term and summations over the different noise terms

$$t_n = n\bar{\tau} + \sum_{i=1}^n (1 - \alpha)^{n-i} \xi_i + \sum_{i=1}^n q_i, \quad (\text{B.23})$$

where

$$\bar{\tau} := \mathbb{E} \frac{\ln(2)}{\kappa_i} \approx \frac{\ln(2)}{\bar{\kappa}} \left(1 + \frac{\sigma_{\kappa}^2}{\bar{\kappa}^2} + 3 \frac{\sigma_{\kappa}^4}{\bar{\kappa}^4} \right) \quad (\text{B.24})$$

and where we defined division time noise variables resulting from growth rate variability as

$$q_n := \frac{\ln(2)}{\kappa_n} - \bar{\tau} \quad (\text{B.25})$$

which are centered around 0 with variance

$$\sigma_q^2 = \mathbb{E} \left(\frac{\ln(2)}{\kappa_n} \right)^2 - \bar{\tau}^2 \approx \frac{\ln(2)^2}{\bar{\kappa}^2} \frac{\sigma_{\kappa}^2}{\bar{\kappa}^2} \approx \bar{\tau}^2 \frac{\sigma_{\kappa}^2}{\bar{\kappa}^2}. \quad (\text{B.26})$$

Now we assume that ξ_n is normal centered around 0 with standard deviation σ_{ξ} , as well as that q_n is normal (which is true when κ_n is normal with small σ_{κ}). We get that t_n are normally distributed with

$$\mathbb{E} t_n = n\bar{\tau} \quad (\text{B.27})$$

and

$$\text{Var}(t_n) = \sum_{i=1}^n (1 - \alpha)^{2n-2i} \sigma_{\xi}^2 + \sum_{i=1}^n \sigma_q^2 \quad (\text{B.28})$$

$$= \sigma_{\Xi_n} + n\sigma_q^2, \quad (\text{B.29})$$

where $\sigma_{\Xi_n} = \sqrt{\frac{1-(1-\alpha)^{2n}}{2\alpha-\alpha^2}}\sigma_\xi$. Plugging this into equation 4.1 yields equation 4.4. One can also calculate the covariance between t_n and t_{n+1} using equation B.22. One can now devise the bivariate Gaussian distribution density function of t_n and t_{n+1} and use this to calculate each of the terms in equation 4.3, namely

$$\frac{d}{d\tau}\mathbb{P}(t-\tau \leq t_n \leq t < t_{n+1}) = \Phi\left(\frac{-\frac{\alpha\sigma_{\Xi_n}^2}{\sigma_{\Xi_n}^2+n\sigma_q^2}(t-n\bar{\tau}-\tau) + (\bar{\tau}-\tau)}{\sqrt{\sigma_\xi^2 + \sigma_q^2 + \frac{n\alpha^2\sigma_q^2\sigma_{\Xi_n}^2}{n\sigma_q^2+\sigma_{\Xi_n}^2}}}\right). \quad (\text{B.30})$$

Plugging this into equation 4.3 yields equation 4.7

B.3 Asymptotic total population

Let us first provide an intuitive sketch behind equation 4.8 in the case of vanishing time-additive noise terms ξ_n . Then later we will prove that this holds for non-vanishing time-additive noise too using theory from Ref. [Fig21] which allows for the calculation of asymptotic growth rate for models with correlated division times.

Consider the case where the time-additive noise terms vanish, We thus have that

$$t_n = \sum_{i=1}^n \tau_i, \quad (\text{B.31})$$

where all division times are independent and given by

$$\tau_n = \frac{\ln(2)}{\kappa_n}. \quad (\text{B.32})$$

Based on the observation that division times desynchronise, one may argue that asymptotically, new cells are born at a speed proportional to the total population. We thus assume that for large t

$$\mathbb{E}N(t) \propto e^{k_\infty t}, \quad (\text{B.33})$$

for some k_∞ that is yet to be determined. At this point one could make the same arguments Powell made as shown in section section 2 to show that κ_∞ must satisfy equation 4.9. We will use this idea but in a more mathematically rigorous sense starting with the general formula we have for the total expected population in equation 4.1. Note that for all n we have that

$$\int \mathbb{P}(t_n \leq t - \tau)g_{\tau_1}(\tau)d\tau = \int \mathbb{P}(t_{n+1} \leq t | \tau_{n+1} = \tau)g_{\tau_1}(\tau)d\tau = \mathbb{P}(t_{n+1} \leq t), \quad (\text{B.34})$$

where we used the fact that all division times τ_n are independent with the same distribution. If we now apply this formula to each term in equation 4.1, we see that

$$2 \int \mathbb{E}N(t - \tau)g_{\tau_1}(\tau)d\tau = 2 + \sum_{n=1}^{\infty} 2^n \int \mathbb{P}(t_n \leq t - \tau)g_{\tau_1}(\tau)d\tau \quad (\text{B.35})$$

$$= 2 + \sum_{n=1}^{\infty} 2^n \mathbb{P}(t_{n+1} \leq t) \quad (\text{B.36})$$

$$= 2 + \sum_{n=2}^{\infty} 2^{n-1} \mathbb{P}(t_n \leq t) \quad (\text{B.37})$$

$$= \mathbb{E}N(t) - \mathbb{P}(\tau_1 \leq t) + 1, \quad (\text{B.38})$$

hence

$$\int \frac{\mathbb{E}N(t - \tau)}{\mathbb{E}N(t)}g_{\tau_1}(\tau)d\tau = \frac{1}{2} - \frac{1 - \mathbb{P}(\tau_1 \leq t)}{2\mathbb{E}N(t)}. \quad (\text{B.39})$$

In the large t limit, the right non-constant term in the right-hand side of the equation vanishes, and we can fill in the assumption $\mathbb{E}N(t) \propto e^{k_\infty t}$ to obtain that

$$\int e^{-k_\infty \tau} g_{\tau_1}(\tau) d\tau = \frac{1}{2}, \quad (\text{B.40})$$

which is Powell's relationship. Through substituting $\tau_1 = \frac{\ln(2)}{\kappa_1}$, one finds equation 4.9. We assumed vanishing time-additive noise and independent division times in this derivation. For correlated division times we can use the generalized Euler-Lotka equation from Ref. [Fig21], which reads

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \left(\mathbb{E} e^{-k_\infty t_n} \right) = -\ln(2) \quad (\text{B.41})$$

in the terminology used in this thesis up to this point. In equation B.22 we derived that approximately

$$t_n = \sum_{i=1}^n (1-\alpha)^{n-i} \xi_i + \sum_{i=1}^n \frac{\ln(2)}{\kappa_i}, \quad (\text{B.42})$$

which can be filled in to give

$$\frac{1}{n} \ln \left(\mathbb{E} e^{-k_\infty t_n} \right) = \frac{1}{n} \ln \left(\mathbb{E} \left[e^{-k_\infty \sum_{i=1}^n (1-\alpha)^{n-i} \xi_i} \right] \right) + \ln \left(\mathbb{E} \left[2^{-\frac{k_\infty}{\kappa_1}} \right] \right). \quad (\text{B.43})$$

For normally distributed or any other realistic noise terms ξ_i for that matter, one can show that $\mathbb{E} \left[e^{-k_\infty \sum_{i=1}^n (1-\alpha)^{n-i} \xi_i} \right]$ converges to some constant value as n goes to infinity, so the first term on the right-hand side must vanish. What we are left with is that equation B.41 turns into

$$\ln \left(\mathbb{E} \left[2^{-\frac{k_\infty}{\kappa_1}} \right] \right) = -\ln(2), \quad (\text{B.44})$$

which can be rewritten as the desired equation

$$\int 2^{-\frac{k_\infty}{\kappa}} \rho(\kappa) d\kappa = \mathbb{E} \left[2^{-\frac{k_\infty}{\kappa_1}} \right] = \frac{1}{2}. \quad (\text{B.45})$$

B.4 Asymptotic age distribution

Here we will try to approximately find out what function equation 4.7 converges to when t goes to infinity. First we note that each term in the summation contains a factor of

$$2^n g_{t_n}(t - \tau) = 2^n \frac{1}{\sqrt{2\pi(\sigma_{\Xi_n}^2 + n\sigma_q^2)}} \exp \left[-\frac{(t - n\bar{\tau} - \tau)^2}{2(\sigma_{\Xi_n}^2 + n\sigma_q^2)} \right], \quad (\text{B.46})$$

where τ is typically no more than $2\bar{\tau}$. One can argue that these terms decay exponentially in the difference between the time and the average number of divisions up to that point, $t - n\bar{\tau}$. So all terms where t and $n\bar{\tau}$ lie too far apart can be ignored. We thus assume that for all terms where $2^n g_{t_n}(t - \tau)$ is not small, it holds that $\frac{t - n\bar{\tau}}{n} \rightarrow 0$ as n and t go to infinity. So for large n and similar t we get that approximately

$$\Phi \left(\frac{-\frac{\alpha\sigma_{\Xi_n}^2}{\sigma_{\Xi_n}^2 + n\sigma_q^2}(t - n\bar{\tau} - \tau) + (\bar{\tau} - \tau)}{\sqrt{\sigma_\xi^2 + \sigma_q^2 + \frac{n\alpha^2\sigma_q^2\sigma_{\Xi_n}^2}{n\sigma_q^2 + \sigma_{\Xi_n}^2}}} \right) \rightarrow \Phi \left(\frac{\bar{\tau} - \tau}{\sqrt{\sigma_\xi^2 + \sigma_q^2 + \alpha^2\sigma_{\Xi_\infty}^2}} \right) \quad (\text{B.47})$$

$$= \Phi \left(\frac{\tau - \bar{\tau}}{\sqrt{\sigma_q^2 + \frac{2}{2-\alpha}\sigma_\xi^2}} \right), \quad (\text{B.48})$$

which is independent of n . If we fill this into equation 4.7, we can factor it out and we obtain that

$$G(t, \tau) = \Phi \left(\frac{\tau - \bar{\tau}}{\sqrt{\sigma_q^2 + \frac{2}{2-\alpha} \sigma_\xi^2}} \right) \frac{1}{\mathbb{E}N(t)} \sum_{i=1}^n 2^i g_{t_n}(t - \tau) \quad (\text{B.49})$$

$$= \Phi \left(\frac{\tau - \bar{\tau}}{\sqrt{\sigma_q^2 + \frac{2}{2-\alpha} \sigma_\xi^2}} \right) \frac{1}{\mathbb{E}N(t)} \sum_{i=1}^n 2^i g_{t_n}(t - \tau) \quad (\text{B.50})$$

$$= \Phi \left(\frac{\tau - \bar{\tau}}{\sqrt{\sigma_q^2 + \frac{2}{2-\alpha} \sigma_\xi^2}} \right) \frac{1}{\mathbb{E}N(t)} 2 \frac{d}{dt} \mathbb{E}N(t - \tau). \quad (\text{B.51})$$

Recall that for large t , we had that $\mathbb{E}N(t)$ becomes an exponential function with rate k_∞ . There thus exists some constant C such that $\mathbb{E}N(t) = Ce^{k_\infty t}$. The asymptotic population growth function now becomes $\frac{d}{dt} \mathbb{E}N(t - \tau) = Ck_\infty e^{k_\infty(t-\tau)}$. Filling this in gives

$$G(t, \tau) = 2k_\infty e^{-k_\infty \tau} \Phi \left(\frac{\tau - \bar{\tau}}{\sqrt{\sigma_q^2 + \frac{2}{2-\alpha} \sigma_\xi^2}} \right). \quad (\text{B.52})$$

C Ornstein-Uhlenbeck growth rate variability model

C.1 Time-dependent formulas

In order to derive the asymptotic size distribution and growth model we will first find the dynamic time dependent size distribution and total population. These functions will look similar to the ones found for the discrete step growth variability model, except instead of comparing the division timings to the global time, we will compare the sizes required to initiate division to the 'global size increase along a lineage'. This may a bit hard to grasp, so let us define the total logarithmic size gain along one lineage after a time t , namely

$$H(t) = \int_0^t \lambda_s ds. \quad (\text{C.1})$$

The idea is that for each division taking place, there is some unique quantity of logarithmic size U_n gained along the lineage of the dividing cell at which this division initiates. This division will have happened if and only if the logarithmic size gained along this lineage at time t exceeds the logarithmic size gained needed to initiate this division, in other words when $U_n \leq H(t)$. We will later explain the definition and derive the distribution of U_n in further detail. First we find the distribution of $H(t)$.

Since λ_s is normally distributed, we have that $H(t)$ is normally distributed too, which has its distribution then fully characterized by its mean and variance, which we will now try to calculate. For this, we only need to know that $\mathbb{E}\lambda_t = \bar{\lambda}$ and $Cov(\lambda_t, \lambda_{t+s}) = \sigma_\lambda^2 e^{-\theta s}$ for all $t, s \geq 0$. We find that

$$\mathbb{E}H(t) = \int_0^t \bar{\lambda} ds = \bar{\lambda} t \quad (\text{C.2})$$

and

$$Var(H(t)) = Cov \left(\int_0^t \lambda_s ds, \int_0^t \lambda_r dr \right) \quad (\text{C.3})$$

$$= \int_0^t \int_0^t Cov(\lambda_s, \lambda_r) dr ds. \quad (\text{C.4})$$

Due to the symmetry of the integrand with respect to switching s and r , integration over the region where $r \leq s$ must have the same outcome as the region where $s \leq r$, so both must be halves of the full integral.

We can therefore integrate over one of the halves and double the outcome to obtain

$$\text{Var}(H(t)) = 2 \int_0^t \int_0^s \text{Cov}(\lambda_s, \lambda_r) dr ds \quad (\text{C.5})$$

$$= 2\sigma_\lambda^2 \int_0^t \int_0^s e^{\theta(r-s)} dr ds \quad (\text{C.6})$$

$$= 2\sigma_\lambda^2 \int_0^t \frac{1}{\theta} (1 - e^{-\theta s}) ds \quad (\text{C.7})$$

$$= 2\frac{\sigma_\lambda^2}{\theta} t - 2\frac{\sigma_\lambda^2}{\theta^2} (1 - e^{-\theta t}). \quad (\text{C.8})$$

Next we want to figure out the total logarithmic size gain U_n needed to initiate an n -th generational division. We can rewrite equation 5.2, which governs the sizes needed per cell to divide, to the logarithmic size recursive relationship

$$u_n = \alpha \bar{u} + (1 - \alpha)u_{n-1} + \eta_n, \quad (\text{C.9})$$

where $u_n = \ln(v_n)$ and $\bar{u} = \ln(\Delta)$. This can be rewritten as

$$\delta u_n = (1 - \alpha)\delta u_{n-1} + \eta_n, \quad (\text{C.10})$$

where $\delta u_n = u_n - \bar{u}$. This recursive relationship is solved for

$$\delta u_n = \sum_{i=1}^n (1 - \alpha)^{n-i} \eta_i. \quad (\text{C.11})$$

In order to find the total logarithmic size gained up till the n -th division, we need to add all the logarithmic sizes gained per generation up to this point. The total logarithmic size increase during the lifetime of an $n-1$ -th generational cell, is the difference between its size at division and its size at birth, which is $u_n + \ln(2) - u_{n-1}$, since the size at division is always twice the birth size of cell next in line. The total logarithmic size at the point of the n -th division is therefore

$$U_n = \sum_{i=1}^n u_i + \ln(2) - u_{n-1} = n \ln(2) + \delta u_n. \quad (\text{C.12})$$

As for the total population in equation C.14, the derivation is similar to what we did before in equation 4.1, except this time we will be comparing total size gained along some lineage instead of elapsed times. Again, we can characterize an n -th generational cell by a binary vector $j = (j_1, \dots, j_{n-1})$. By $H^j(t)$ we denote the total logarithmic size increase along the cell lineage $(j_1, \dots, j_{n-1}, 0, 0, \dots)$ at time t . Let U_n^j be the total logarithmic size gained along the lineage of cell j at the the birth of cell j . An n -th generational cell j is therefore present if and only if $U_n^j \leq H^j(t) < U_{n+1}^j$. The total number of cells present at time t can be given by

$$N(t) = \sum_{j \in \{0,1\}^n} \mathbf{1}\{U_n^j \leq H^j(t) < U_{n+1}^j\} = 1 + \sum_{n=1}^{\infty} \sum_{j \in \{0,1\}^{n-1}} \mathbf{1}\{U_n^j \leq H^j(t)\}, \quad (\text{C.13})$$

which, by exploiting the linearity of taking expectation values and the fact that each lineage has the same distribution functions, results in

$$\mathbb{E}N(t) = 1 + \sum_{n=1}^{\infty} 2^{n-1} \mathbb{P}(n \ln(2) + \delta u_n \leq H(t)). \quad (\text{C.14})$$

Under the assumption that the division size noise terms η_n are normally distributed around 0 with standard deviation σ_η , we find that using equation C.11, the birth size deviations δu_n are normally distributed around 0 with variance

$$\text{Var}(\delta u_n) = \sum_{i=1}^n (1 - \alpha)^{2n-2i} \sigma_\eta^2 = \frac{1 - (1 - \alpha)^{2n}}{2\alpha - \alpha^2} \sigma_\eta^2. \quad (\text{C.15})$$

Filling this into the general expected population formula, equation C.14 yields

$$\mathbb{E}N(t) = 1 + \sum_{n=1}^{\infty} 2^{n-1} \Phi \left(\frac{\bar{\lambda}t - n \ln(2)}{\sqrt{2 \frac{\sigma_{\lambda}^2}{\theta} t - 2 \frac{\sigma_{\lambda}^2}{\theta^2} (1 - e^{-\theta t})} + \frac{1 - (1-\alpha)^{2n}}{\alpha(2-\alpha)} \sigma_{\eta}^2}} \right). \quad (\text{C.16})$$

Next we wish to find the size distribution. The approach is similar to what we used in the derivation of the age distribution in the discrete step growth rate model. Suppose an n -th generational cell j is present at time t . Assuming that the birth size of the first cell is the equilibrium size, the logarithmic size of j is the difference between the total size gained along this lineage up to time t and the total size lost due to divisions. Since each division splits the size exactly in half, the logarithmic size lost must be precisely $n \ln(2)$. The current logarithmic size increase with respect to the equilibrium size is thus $H^j(t) - n \ln(2)$. A cell j thus has a size smaller than v if and only if $H^j(t) - n \ln(2) \leq \ln(v/\Delta)$, hence the total number of cells at time t currently smaller than v is

$$T(t, v) = \sum_{n=0}^{\infty} \sum_{j \in \{0,1\}^n} \mathbb{1}\{H^j(t) - n \ln(2) \leq \ln(v/\Delta), U_n^j \leq H^j(t) < U_{n+1}^j\}, \quad (\text{C.17})$$

which has an expectation of

$$\mathbb{E}T(t, v) = \sum_{n=0}^{\infty} 2^n \mathbb{P}(H(t) - n \ln(2) \leq \ln(v/\Delta), \delta u_n \leq H^j(t) - n \ln(2) < \ln(2) + \delta u_{n+1}). \quad (\text{C.18})$$

We can find the size density by taking the derivative with respect to size and dividing it by the total population

$$G(t, v) = \frac{d}{dv} \frac{\mathbb{E}T(t, v)}{\mathbb{E}N(t)} \quad (\text{C.19})$$

$$= \sum_{n=0}^{\infty} 2^n \frac{d}{dv} \mathbb{P}(H(t) - n \ln(2) \leq \ln(v/\Delta), \delta u_n \leq H(t) - n \ln(2) < \ln(2) + \delta u_{n+1}). \quad (\text{C.20})$$

Since $H(t)$ and δu_n are independent, taking the derivative with respect to the probability in the above summation results in

$$\frac{d}{dv} \mathbb{P}(H(t) - n \ln(2) \leq \ln(v/\Delta), \delta u_n \leq H(t) - n \ln(2) < \ln(2) + \delta u_{n+1}) \quad (\text{C.21})$$

$$= \frac{1}{v} g_{H(t)}(\ln(v/\Delta) + n \ln(2)) \mathbb{P}(\delta u_n \leq \ln(v/\Delta) < \ln(2) + \delta u_{n+1}), \quad (\text{C.22})$$

where $g_{H(t)}$ is the density distribution function of $H(t)$. Physically its impossible and mathematically extremely unlikely to have that the birth size of some cell is larger than its size at division, so we can assume that $\delta u_n \leq \ln(2) + \delta u_{n+1}$ with full probability. This lets us rewrite the above expression as

$$\frac{1}{v} g_{H(t)}(\ln(v/\Delta) + n \ln(2)) (\mathbb{P}(\delta u_n \leq \ln(v/\Delta) - \mathbb{P}(\delta u_{n+1} \leq \ln(v/\Delta) - \ln(2))), \quad (\text{C.23})$$

hence

$$G(t, v) = \frac{1}{v} \sum_{n=0}^{\infty} 2^n g_{H(t)}(\ln(v/\Delta) + n \ln(2)) (\mathbb{P}(\delta u_n \leq \ln(v/\Delta) - \mathbb{P}(\delta u_{n+1} \leq \ln(v/\Delta) - \ln(2))). \quad (\text{C.24})$$

If we again assume that η_n are normally distributed around 0 with standard deviation σ_{η} , then we can greatly simplify this equation and turn it into a more usable expression, namely

$$G(t, v) = \frac{1}{v \mathbb{E}N(t)} \sum_{n=1}^{\infty} g_{H(t)}(\ln(v/\Delta) + n \ln(2)) \left(\Phi \left(\frac{\ln(v/\Delta)}{\sqrt{\frac{1 - (1-\alpha)^{2n}}{\alpha(2-\alpha)} \sigma_{\eta}^2}} \right) - \Phi \left(\frac{\ln(v/\Delta) - \ln(2)}{\sqrt{\frac{1 - (1-\alpha)^{2(n+1)}}{\alpha(2-\alpha)} \sigma_{\eta}^2}} \right) \right), \quad (\text{C.25})$$

where

$$g_{H(t)}(\ln(v/\Delta) + n \ln(2)) = \frac{1}{\sqrt{2\pi(2 \frac{\sigma_{\lambda}^2}{\theta} t - 2 \frac{\sigma_{\lambda}^2}{\theta^2} (1 - e^{-\theta t}))}} \exp \left[-\frac{(\ln(v/\Delta) + n \ln(2) - \bar{\lambda}t)^2}{2(2 \frac{\sigma_{\lambda}^2}{\theta} t - 2 \frac{\sigma_{\lambda}^2}{\theta^2} (1 - e^{-\theta t}))} \right]. \quad (\text{C.26})$$

C.2 Asymptotic total population

The goal of this section is to find the asymptotic growth rate of equation C.14, in a similar vein to what we did for the discrete step growth rate variability model. We first define

$$\mathbb{E}N(t, x) = 1 + \sum_{n=1}^{\infty} 2^{n-1} \mathbb{P}(n \ln(2) + \delta u_n \leq H(t) + x), \quad (\text{C.27})$$

which counts the expected number of divisions that happened at an accumulated lineage logarithmic size gain of less than $H(t) + x$. We now note that

$$\mathbb{E}N(t, x + \ln(2)) = 1 + \sum_{n=1}^{\infty} 2^{n-1} \mathbb{P}((n-1) \ln(2) + \delta u_n \leq H(t) + x) \quad (\text{C.28})$$

$$= 2 \left(1 + \sum_{n=1}^{\infty} 2^{n-1} \mathbb{P}(n \ln(2) + \delta u_{n+1} \leq H(t) + x) \right). \quad (\text{C.29})$$

For large t , the terms with small n become negligible with respect to terms with large n . For large n we found that the birth size deviation quickly converges, so we must also have that $\delta u_{n+1} \sim \delta u_n$ asymptotically as n becomes large. As a result equation C.28 turns into

$$\mathbb{E}N(t, x + \ln(2)) = 2\mathbb{E}N(t, x). \quad (\text{C.30})$$

We now assume that $\mathbb{E}(t, x)$ for large t and finite x is not just exponential in t but also in x . Equation C.30 tells us that the population doubles every time we increase x by $\ln(2)$, hence we may write

$$\mathbb{E}N(t, x) = e^x \mathbb{E}N(t). \quad (\text{C.31})$$

Recall that for large t we have that

$$H(t) \sim \text{Normal} \left(\bar{\lambda}t, 2 \frac{\sigma_\lambda^2}{\theta} \left(t - \frac{1}{\theta} \right) \right). \quad (\text{C.32})$$

Let us now also define another process H' that is independent of H and satisfies

$$H'(t) \sim \text{Normal} \left(\bar{\lambda}t, 2 \frac{\sigma_\lambda^2}{\theta} t \right). \quad (\text{C.33})$$

By comparing means and variances, one finds that for all s

$$H(t) + H'(s) \sim H(t + s). \quad (\text{C.34})$$

It follows that

$$\int \mathbb{P}(n \ln(2) + \delta u_n \leq H(t) + x) g_{H'(s)}(x) dx = \mathbb{P}(n \ln(2) + \delta u_n \leq H(t + s)). \quad (\text{C.35})$$

Applying this relation to every term in equation C.27 results in

$$\int \mathbb{E}N(t, x) g_{H'(s)}(x) dx = \mathbb{E}N(t + s). \quad (\text{C.36})$$

Plugging equation C.31 into this gives

$$\mathbb{E}N(t + s) = \int e^x \mathbb{E}N(t) g_{H'(s)}(x) dx \quad (\text{C.37})$$

$$= \mathbb{E}N(t) \int e^x \frac{1}{\sqrt{4\pi \frac{\sigma_\lambda^2}{\theta} s}} \exp \left[-\frac{(x - \bar{\lambda}s)^2}{4 \frac{\sigma_\lambda^2}{\theta} s} \right] dx \quad (\text{C.38})$$

$$= \mathbb{E}N(t) \int \frac{1}{\sqrt{4\pi \frac{\sigma_\lambda^2}{\theta} s}} \exp \left[-\frac{(x - 2 \frac{\theta}{\sigma_\lambda^2} s - \bar{\lambda}s)^2}{4 \frac{\sigma_\lambda^2}{\theta} s} + \frac{\sigma_\lambda^2}{\theta} s + \bar{\lambda}s \right] dx \quad (\text{C.39})$$

$$= \mathbb{E}N(t) \exp \left[\left(\bar{\lambda} + \frac{\sigma_\lambda^2}{\theta} \right) s \right], \quad (\text{C.40})$$

hence we have that asymptotically

$$\mathbb{E}N(t) \propto e^{\lambda_\infty t} \quad (\text{C.41})$$

where

$$\lambda_\infty = \bar{\lambda} + \frac{\sigma_\lambda^2}{\theta}. \quad (\text{C.42})$$

C.3 Asymptotic size distribution

We now wish to find what equation C.24 converges to for large t . When t is large we can argue that low n terms become negligible and for high n terms, the distributions of δu_n quickly converge to some distribution satisfying

$$\delta u_\infty \sim \sum_{i=1}^{\infty} (1-\alpha)^{i-1} \eta_i, \quad (\text{C.43})$$

hence asymptotically we get

$$G(t, v) \quad (\text{C.44})$$

$$= \frac{1}{v\mathbb{E}N(t)} \sum_{n=1}^{\infty} 2^n g_{H(t)}(\ln(v/\Delta) + n \ln(2)) (\mathbb{P}(\delta u_\infty \leq \ln(v/\Delta)) - \mathbb{P}(\delta u_\infty \leq \ln(v/\Delta) - \ln(2))) \quad (\text{C.45})$$

$$= \mathbb{P}(\ln(v/\Delta) - \ln(2) \leq \delta u_\infty \leq \ln(v/\Delta)) \frac{1}{v\mathbb{E}N(t)} \sum_{n=1}^{\infty} 2^n g_{H(t)}(\ln(v/\Delta) + n \ln(2)). \quad (\text{C.46})$$

Let us focus on the infinite summation here. Note that for each term we can write

$$g_{H(t)}(\ln(v/\Delta) + n \ln(2)) = -v \frac{d}{dv} \mathbb{P}(\ln(v/\Delta) + n \ln(2) \leq H(t)), \quad (\text{C.47})$$

hence the infinite summation becomes

$$\sum_{n=1}^{\infty} 2^n g_{H(t)}(\ln(v/\Delta) + n \ln(2)) = -v \frac{d}{dv} \sum_{n=1}^{\infty} 2^n \mathbb{P}(n \ln(2) \leq H(t) - \ln(v/\Delta)) \quad (\text{C.48})$$

$$= -v \frac{d}{dv} 2\mathbb{E}N_{\eta=0}(t, -\ln(v/\Delta)), \quad (\text{C.49})$$

where $\mathbb{E}N_{\eta=0}(t, x)$ corresponds to equation C.27 but with the birth size noise parameters η_n and thus any birth size deviations δu_n set to zero. Through an analogous derivation or by noting that the asymptotic population growth is independent of birth size noise, we find that

$$\mathbb{E}N_{\eta=0}(t, x) \propto e^x e^{\lambda_\infty t}, \quad (\text{C.50})$$

hence

$$\sum_{n=1}^{\infty} 2^n g_{H(t)}(\ln(v/\Delta) + n \ln(2)) \propto -v \frac{d}{dv} \frac{\Delta}{v} e^{\lambda_\infty t} = \frac{\Delta}{v} e^{\lambda_\infty t}, \quad (\text{C.51})$$

which can be filled back in equation C.46 to give

$$G(v) \propto \frac{\Delta}{v^2} \mathbb{P}(\ln(v/\Delta) - \ln(2) \leq \delta u_\infty \leq \ln(v/\Delta)). \quad (\text{C.52})$$

All that remains is for us to find the right normalization constant C by calculating

$$1 = C \int_0^\infty \frac{\Delta}{v^2} \mathbb{P}(\ln(v/\Delta) - \ln(2) \leq \delta u_\infty \leq \ln(v/\Delta)) dv \quad (\text{C.53})$$

$$= C \int_0^\infty \int_{-\infty}^\infty \frac{\Delta}{v^2} \mathbb{1}\{\ln(v/\Delta) - \ln(2) \leq \delta u \leq \ln(v/\Delta)\} g_{\delta u_\infty}(\delta u) d\delta u dv \quad (\text{C.54})$$

$$= C \int_{-\infty}^\infty \int_0^\infty \frac{\Delta}{v^2} \mathbb{1}\{\Delta e^{\delta u} \leq v \leq 2\Delta e^{\delta u}\} g_{\delta u_\infty}(\delta u) dv d\delta u \quad (\text{C.55})$$

$$= C \int_{-\infty}^\infty \left(\frac{\Delta}{2\Delta e^{\delta u}} - \frac{\Delta}{\Delta e^{\delta u}} \right) g_{\delta u_\infty}(\delta u) d\delta u \quad (\text{C.56})$$

$$= \frac{C}{2} \int_{-\infty}^\infty e^{-\delta u} g_{\delta u_\infty}(\delta u) d\delta u, \quad (\text{C.57})$$

which means that $C = \frac{2}{\mathbb{E}e^{-\delta u_\infty}}$ thus resulting in equation 6.5.

D Comparison of discrete-step growth rate and Ornstein-Uhlenbeck growth rate

In this section we are interested in comparing the different models and specifically their asymptotic population growth rates. Since in neither model the direct division time or cell division size noise affects the population growth rate, we will ignore them. The discrete-step growth rate κ_1 should correspond to the time-average growth rate of a first generation cell, hence we can relate such a variable to the growth rate process via

$$\kappa_1 = \frac{1}{\tau_1} \int_0^{\tau_1} \lambda_s ds \quad (\text{D.1})$$

where τ_1 is the timing of the first division. Since there is no division size noise, we know that the cell must have doubled in size at the time of the first division, hence τ_1 must also satisfy

$$H(\tau_1) = \int_0^{\tau_1} \lambda_s ds = \ln(2). \quad (\text{D.2})$$

Physically, one can expect the growth rate to stay positive at all times making $H(\tau)$ an increasing stochastic function of τ , so there should be no difference distribution-wise between defining τ_1 as the first hitting time satisfying

$$\tau_1 = \inf\{t \geq 0 : H(t) = \ln(2)\} \quad (\text{D.3})$$

and as a random variable satisfying

$$\mathbb{P}(\tau_1 \leq \tau) = \mathbb{P}(\ln(2) \leq H(\tau)). \quad (\text{D.4})$$

Unfortunately, there is a mathematical possibility of the growth rate being negative. As we shall later see, the difference between the average cell average growth rate κ when defining τ_1 as a first hitting time and defining it through equation D.2 is of the same order as the difference between the asymptotic growth rate and the time average cell growth rate, so it cannot be assumed to be negligible. To demonstrate this difference, we will consider the special case where $\frac{\rho}{\lambda} \gg 1$ while $D := \frac{\sigma_\lambda^2}{\theta}$ is kept constant, which is the case of vanishing auto-correlation. The integral defined in equation ?? is now a normally distributed function with $\mathbb{E}H(t) = \bar{\lambda}t$ and $\text{Var}(H(t)) = 2Dt$. Furthermore, one can show that $\text{Cov}(H(t+s) - H(t), H(t)) = 0$ for all $t, s \geq 0$, so $H(t)$ is a Brownian motion process with drift term $\bar{\lambda}$ and diffusion constant D . Finding the distribution of the first hitting time of a Brownian motion with drift is a known solvable problem. The probability density function of τ_1 given in equation D.3 is

$$g_{\tau_1}(\tau) = \frac{\ln(2)}{\sqrt{4\pi D\tau^3}} \exp\left[-\frac{(\bar{\lambda}\tau - \ln(2))^2}{4D\tau}\right]. \quad (\text{D.5})$$

Let us define $\bar{\tau} = \ln(2)/\bar{\lambda}$ and $d\tau = \tau - \bar{\lambda}$. We will now factor out a normal density distribution function and expand the prefactor up to first order in $\frac{d\tau}{\bar{\tau}} \sim \frac{\sqrt{D\bar{\tau}}}{\lambda}$

$$g_{d\tau_1}(d\tau) = \frac{\ln(2)}{\sqrt{4\pi D(\bar{\tau} + d\tau)^3}} \exp\left[-\frac{\bar{\lambda}^2 d\tau^2}{4D(\bar{\tau} + d\tau)}\right] \quad (\text{D.6})$$

$$\approx \frac{\bar{\lambda}}{\sqrt{4\pi D\bar{\tau}}} \left(1 - \frac{3}{2} \frac{1}{\bar{\tau}} d\tau\right) \left(1 + \frac{\bar{\lambda}^2}{4D\bar{\tau}^2} d\tau^3\right) \exp\left[-\frac{\bar{\lambda}^2 d\tau^2}{4D\bar{\tau}}\right] \quad (\text{D.7})$$

$$= \left(1 - \frac{3}{2\bar{\tau}} d\tau + \frac{\bar{\lambda}^2}{4D\bar{\tau}^2} d\tau^3\right) g_{d\tau_1}^*(d\tau), \quad (\text{D.8})$$

where $g_{d\tau_1}^*$ is the density distribution function of a centered normal distribution with variance $\mathbb{E}^*(d\tau_1^2) = \frac{2D\bar{\tau}}{\bar{\lambda}^2}$. By \mathbb{E}^* we denote the expectation value taken over $g_{d\tau_1}^*$ instead of $g_{d\tau_1}$. By combining equation D.1 and D.2 we obtain that

$$\bar{\kappa} = \mathbb{E}\kappa_1 = \mathbb{E} \frac{\ln(2)}{\tau_1}. \quad (\text{D.9})$$

We will solve this up to second order in $\frac{d\tau}{\bar{\tau}} \sim \frac{\sqrt{D\bar{\tau}}}{\lambda}$, so

$$\bar{\kappa} = \frac{\ln(2)}{\bar{\tau}} \mathbb{E} \left[\frac{1}{1 + \frac{d\tau_1}{\bar{\tau}}} \right] \quad (\text{D.10})$$

$$= \bar{\lambda} \mathbb{E} \left[1 - \frac{1}{\bar{\tau}} d\tau_1 + \frac{1}{\bar{\tau}^2} d\tau_1^2 \right] \quad (\text{D.11})$$

$$= \bar{\lambda} \mathbb{E}^* \left[\left(1 - \frac{1}{\bar{\tau}} d\tau_1 + \frac{1}{\bar{\tau}^2} d\tau_1^2\right) \left(1 - \frac{3}{2\bar{\tau}} d\tau_1 + \frac{\bar{\lambda}^2}{4D\bar{\tau}^2} d\tau_1^3\right) \right]. \quad (\text{D.12})$$

Due to the symmetry of $g_{d\tau_1}^*$ all odd terms in $d\tau_1$ vanish under the expectation and we keep

$$\bar{\kappa} = \bar{\lambda} \mathbb{E}^* \left[1 + \frac{5}{2\bar{\tau}} d\tau_1^2 - \frac{\bar{\lambda}^2}{4D\bar{\tau}^3} d\tau_1^4 \right] \quad (\text{D.13})$$

$$= \bar{\lambda} \left(1 + \frac{5}{2\bar{\tau}^2} \frac{2D\bar{\tau}}{\bar{\lambda}^2} - \frac{\bar{\lambda}^2}{4D\bar{\tau}^3} 3 \left(\frac{2D\bar{\tau}}{\bar{\lambda}^2} \right)^2 \right) \quad (\text{D.14})$$

$$= \bar{\lambda} + \frac{2}{\ln(2)} D \quad (\text{D.15})$$

$$= \bar{\lambda} + \frac{2}{\ln(2)} \frac{\sigma_\lambda^2}{\theta}. \quad (\text{D.16})$$

In a similar fashion we can calculate the discrete cell average growth rate variance as

$$\sigma_\kappa^2 = \mathbb{E}\kappa_1^2 - \bar{\kappa}^2 \quad (\text{D.17})$$

$$= \bar{\lambda} \frac{2}{\ln(2)} \frac{\sigma_\lambda^2}{\theta}. \quad (\text{D.18})$$

We can now plug these equations into equation 4.10, which is the formula for the discrete growth rate model and we obtain up to first order in $\frac{\sigma_\lambda^2}{\theta}$ that

$$k_\infty = \bar{\kappa} - \left(1 - \frac{\ln(2)}{2}\right) \frac{\sigma_\kappa^2}{\bar{\kappa}} \quad (\text{D.19})$$

$$= \bar{\lambda} + \frac{2}{\ln(2)} \frac{\sigma_\lambda^2}{\theta} - \left(1 - \frac{\ln(2)}{2}\right) \frac{2}{\ln(2)} \frac{\sigma_\lambda^2}{\theta} \quad (\text{D.20})$$

$$= \bar{\lambda} + \frac{\sigma_\lambda^2}{\theta}. \quad (\text{D.21})$$

which is in perfect agreement with equation 6.2. So to sum things up, the discrete uncorrelated growth rate model and the Ornstein-Uhlenbeck growth rate model with vanishing auto-correlation produce the same asymptotic growth rate up to first order corrections, as long as you account for the difference between the global time average growth rate and the per-cell time average growth rate. This is if we use equation D.3 as our definition of τ_1 . If we used equation D.4 instead, we would find a different result. The division time density function could now be obtained by simply taking the derivative of equation D.4 with respect to τ , which is

$$g_{\tau_1}(\tau) = \frac{d}{d\tau} \Phi\left(\frac{\bar{\lambda}\tau - \ln(2)}{\sqrt{2D\tau}}\right) = \left(\frac{\bar{\lambda}}{\sqrt{4\pi D\tau}} - \frac{1}{2} \frac{\bar{\lambda}\tau - \ln(2)}{\sqrt{4\pi D\tau^3}}\right) \exp\left[-\frac{(\bar{\lambda}\tau - \ln(2))^2}{2D\tau}\right]. \quad (\text{D.22})$$

Through an analogous derivation one finds that using this density we get a slightly different average per-cell time average growth rate, namely

$$\bar{\kappa} = \bar{\lambda} + \frac{1}{\ln(2)} \frac{\sigma_{\bar{\lambda}}^2}{\theta} \quad (\text{D.23})$$

and the same per-cell growth rate variance $\sigma_{\bar{\kappa}}^2$. This results in a different expression of the asymptotic growth rate derived from equation 4.10, namely

$$k_{\infty} = \bar{\lambda} - \left(\frac{1}{\ln(2)} - 1\right) \frac{\sigma_{\bar{\lambda}}^2}{\theta}. \quad (\text{D.24})$$

A numerical check also confirms that the 'wrong' asymptotic growth rate from equation D.24 matches the constant in Powell's relationship, which is equation 2.1, given the 'wrong' division time density from equation D.22. The interesting thing is that equation C.14, which is the formula we derived for the expected total population uses the 'wrong' type of hitting times like in equation D.4, yet both numerical and theoretical checks show that its asymptotic growth rate matches the 'right' formula given in equation 6.2. An explanation for this discrepancy may lie in the fact that the 'wrong' cumulative total division time probabilities in the summation of C.14 converge to the 'right' probabilities as the generation number n goes to infinity. It is also worth noting that the 'right' first hitting times defined in equation D.3 are mathematically independent of one another in the Brownian motion limit whereas the 'wrong' hitting times in equation D.4 are not perfectly independent. Powell's relationship and by extension, equation 4.10 assume independence of division times in their derivations. What is most unfortunate is that the 'right' first hitting time density formula can only be derived in the Brownian motion limit, since there is no known analytical formula for the proper first hitting time of a general Ornstein-Uhlenbeck process [BS12; LK18].

References

- [Doo42] Joseph L Doob. "The Brownian movement and stochastic equations". In: *Annals of Mathematics* (1942), pp. 351–369.
- [Pow56] EO Powell. "Growth rate and generation time of bacteria, with special reference to continuous culture". In: *Microbiology* 15.3 (1956), pp. 492–511.
- [KMS58] NO Kjeldgaard, O Maaløe, and Moselio Schaechter. "The transition between different physiological states during balanced growth of *Salmonella typhimurium*". In: *Microbiology* 19.3 (1958), pp. 607–616.
- [SMK58] Moselio Schaechter, Ole Maaløe, and Niels O Kjeldgaard. "Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*". In: *Microbiology* 19.3 (1958), pp. 592–606.
- [TW80] Frank J Trueba and Conrad L Woldringh. "Changes in cell diameter during the division cycle of *Escherichia coli*." In: *Journal of bacteriology* 142.3 (1980), pp. 869–878.
- [Ste+05] Eric J Stewart et al. "Aging and death in an organism that reproduces by morphologically symmetric division". In: *PLoS Biol* 3.2 (2005), e45.
- [Wan+10] Ping Wang et al. "Robust growth of *Escherichia coli*". In: *Current biology* 20.12 (2010), pp. 1099–1103.

- [BS12] Andrei N Borodin and Paavo Salminen. *Handbook of Brownian motion-facts and formulae*. Birkhäuser, 2012.
- [Ami14] Ariel Amir. “Cell size regulation in bacteria”. In: *Physical review letters* 112.20 (2014), p. 208102.
- [Cam+14] Manuel Campos et al. “A constant size extension drives bacterial cell size homeostasis”. In: *Cell* 159.6 (2014), pp. 1433–1446.
- [Rob+14] Lydia Robert et al. “Division in *Escherichia coli* is triggered by a size-sensing rather than a timing mechanism”. In: *BMC biology* 12.1 (2014), pp. 1–10.
- [Yos+14] Mari Yoshida et al. “Directed evolution of cell size in *Escherichia coli*”. In: *BMC evolutionary biology* 14.1 (2014), pp. 1–12.
- [Tah+15] Sattar Taheri-Araghi et al. “Cell-size control and homeostasis in bacteria”. In: *Current biology* 25.3 (2015), pp. 385–391.
- [Has+16] Mikihiro Hashimoto et al. “Noise-driven growth rate gain in clonal cellular populations”. In: *Proceedings of the National Academy of Sciences* 113.12 (2016), pp. 3251–3256.
- [FCH17] Giuseppe Facchetti, Fred Chang, and Martin Howard. “Controlling cell size through sizer mechanisms”. In: *Current Opinion in Systems Biology* 5 (2017), pp. 86–92.
- [LA17] Jie Lin and Ariel Amir. “The effects of stochasticity at the single-cell level and cell size control on the population growth”. In: *Cell systems* 5.4 (2017), pp. 358–367.
- [PJI17] Dan Pirjol, Farshid Jafarpour, and Srividya Iyer-Biswas. “Phenomenology of stochastic exponential growth”. In: *Physical Review E* 95.6 (2017), p. 062406.
- [Jaf+18] Farshid Jafarpour et al. “Bridging the timescales of single-cell and population dynamics”. In: *Physical Review X* 8.2 (2018), p. 021007.
- [Jun+18] Suckjoon Jun et al. “Fundamental principles in bacterial physiology—history, recent progress, and the future with focus on cell size control: a review”. In: *Reports on Progress in Physics* 81.5 (2018), p. 056601.
- [LK18] Alexander Lipton and Vadim Kaushansky. “On the first hitting time density of an Ornstein-Uhlenbeck process”. In: *arXiv preprint arXiv:1810.02390* (2018).
- [Jaf19] Farshid Jafarpour. “Cell size regulation induces sustained oscillations in the population growth rate”. In: *Physical Review Letters* 122.11 (2019), p. 118101.
- [LA20] Jie Lin and Ariel Amir. “From single-cell variability to population growth”. In: *Physical Review E* 101.1 (2020), p. 012401.
- [Pig21] Simone Pigolotti. “Generalized Euler-Lotka equation for correlated cell divisions”. In: *arXiv preprint arXiv:2101.02403* (2021).