MSc Mathematics

Track: Biomedical Mathematics

Master thesis

Stochastics of growing and persisting cell populations

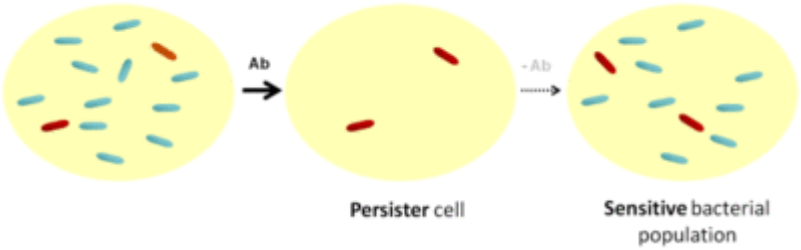
by

Tom van den Bosch

August 25, 2017

Supervisor: dr. Bob Planqué

Second examiner: prof. dr. Frank J. Bruggeman



Department of Mathematics
Faculty of Sciences

VU UNIVERSITY AMSTERDAM

Abstract

Some cellular populations contain subpopulations which exhibit a phenotypic switch to become dormant. These persister cells do not grow, but are able to survive outside stress to reduce the risk of total populations extinction. However, the rate at which cells switch their phenotype is random, and as such the number of persisters shows stochasticity. Because of this, there is a nonzero probability that a population contains no persisters, thus being at risk. In this paper we aim to study this risk. We first derive a master equation for the persister cell model. Since fractions of persister cells are usually low ($10^{-6} - 10^{-5}$), the persister model is very similar to the exponential growth process, for which we derive an exact probability distribution. Using this, we derive the probability generating function of a simplified model for persisters, which we show to be accurate for low fractions. We then fit distributions to data generated by the Gillespie algorithm to show the distribution of persister cells. Lastly, we solve the master equation for the first and second moments to derive expressions for the fraction of persister cells and the noise in the number of cells. With this fraction we derive an exact probability distribution for a model in which the fraction is constant.

Keywords: Persister cells, master equation, probability generating function, branching process, Gillespie algorithm

Acknowledgements

First and foremost, I would like to thank my supervisors Bob Planqué and Frank Bruggeman for their help with this project, but also for their courses in Mathematical Biology that convinced me to take on this subject, for without them I would not have been in this field. More specifically, I would like to thank Frank for helping me understand the questions that I needed to answer, which was of great importance to writing a proper thesis. I would like to thank Bob for helping me find solutions to these problems in ways that I did not think of, and I am convinced that because of this I was able to write my thesis at a mathematical level that I am very content with.

Second, I would like to thank the people in the O2 first floor, as I loved working there. Most notably I would like to thank the other students at our workspace Stefan, Ivar, Gius, Michael, Maurice, Patrick, Mirushe, Camen and Paula for making my days more enjoyable.

Contents

1. Introduction

1.1. Biological relevance

A challenge that emerged with advances in modern medicine is the phenomenon of bacterial persistence. Bacterial persistence is caused by a cell changing its phenotype to go into a 'dormant' state in which it does not grow, but shows multidrug tolerance [1][2]. It was first discovered in 1944 by J. W. Bigger [3] as a cause of reduced success of penicillin therapy for treating staphylococcal infections. However, due to low fractions ($\approx 10^{-6} - 10^{-5}$) of persister cells, isolating sufficient numbers for experimental research proved difficult. This improved when Moyes [4] was able to isolate populations of mutants called Hip (high induced persistence) which contained persister fractions as high as 10^{-2}. Although since then multiple mechanisms for this switch have been found [5][6][7][8], the exact workings are still unclear [9][10][11]. Specifically, little is known about the mechanisms behind the reverse switch from the dormant state to the growing state [12].

Closely related to persistence is antimicrobial resistance, which is associated with a mutation that causes a cell and its offspring to be permanently resistant to drugs. However, this resistance often comes at the cost of reduced fitness in the absence of these drugs. This loss of fitness is much less prevalent in populations containing persister cells [13], as persistence is caused by a phenotypic switch as opposed to mutation. Moreover, persistence is a reversible process in which persister cells that reverse their phenotypic switch are indistinguishable from other normal cells.

Within persister cells, we distinguish between type I and type II persisters [1][2]. Type I persisters are triggered into the dormant state by external conditions such as antibiotics or a lack of nutrients. When not exposed to stressful conditions, these cells resist persistence. Type II persisters are continually produced without environmental triggers, and as such these populations always contain a subpopulation of dormant cells. Because of inherent stochastics, a population of type II persisters will always have a positive probability of containing zero persister cells. As such, bacteria have to make a tradeoff between a lower probability of zero persisters - which means a higher average number of persister cells - and a higher growth rate in environments that lack stress. This causes their behaviour to be similar to bet-hedging strategies [2], which we will not go in to detail in this paper but can be found elsewhere [14][15].

As with most biological processes, growth rates and switching rates in persister cell populations are not constants but rather random variables. For this thesis, we assume all waiting times to be exponentially distributed random variables, leading to mass-action kinetics associated with for instance chemical reactions. The reason for this assumption

is that this is the only (continuous) distribution which has the 'memoryless' property (see Appendix). This means that the future is dependent only on the present, which makes the process of cell growth a Markov process, with ordinary differential equations as its 'macroscopic' desciption, and the chemical master equation as its mesoscopic desciption [16][17]. Although this assumption might not seem biologically reasonable at first, for large populations of cells the age distribution is uniform, making the time until the next event an exponential random variable. Aside from this, markov processes are mathematically much simpler to study than non-Markovian processes, and numerical solutions for Markov processes are also much less computationally intensive than their counterpart. In the next section we construct a mathematical model for the Markov process associated with persister growth and phenotype switching.

1.2. Mathematical model and the master equation

Consider a population of growing cells of integer size g and persister cells of size p. Both grow with rates μ_g and μ_p respectively. Although μ_p is often close to 0, for the sake of the model it is left in. The rates at which cells switch between populations are v_1 and v_2, as seen in Figure 1.1.

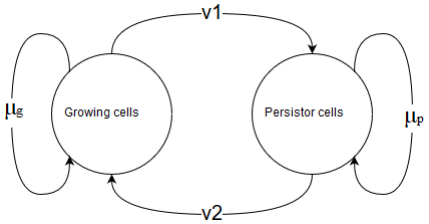


Figure 1.1.: A model for a population of growing and persister cells

Or in equations:

$$g \xrightarrow{\mu_g} 2g, \qquad p \xrightarrow{\mu_p} 2p, \qquad g \underset{v_2}{\overset{v_1}{\rightleftarrows}} p. \tag{1.1}$$

The probability of finding g growing and p persister cells at time t is denoted by $P_{g,p}(t)$. Assuming that all waiting times are exponentially distributed, the process is a Markov process (see A.1). Thus we get a master equation for $P_{g,p}$:

$$\begin{aligned}\frac{d}{dt}P_{g,p} = \mu_g[(g-1)P_{g-1,p} - gP_{g,p}] + \mu_p[(p-1)P_{g,p-1} - pP_{g,p}] \\ + v_1[(g+1)P_{g+1,p-1} - gP_{g,p}] + v_2[(p+1)P_{g-1,p+1} - pP_{g,p}].\end{aligned} \tag{1.2}$$

A derivation of this is given in A.2. Aside from the master equation, we also study the probability generating function (pgf) of $P_{g,p}$ which is defined as

$$F(s_1, s_2, t) = \sum_{s_1, s_2 \in \mathbb{Z}_{\geq 0}} s_1^g s_2^p P_{g,p}(t). \tag{1.3}$$

Note that the sum is over all nonnegative integers in both s_1 and s_2. This generating function will prove useful in mathematically analyzing systems like the persister cell process, as will be seen in later sections. An extremely helpful property of the generating function is that if the generating function is known, the full probability distribution is also known, as $P_{g,p}(t) = \frac{\partial^{g+p}F}{\partial s_1^g \partial s_2^p}(0,0,t)$. Thus, finding the pgf is enough to solve the entire model!

1.3. Research goals

The main reason for the existence of persister cells is for populations to survive stress, and as such a population containing zero persisters is at risk of being wiped out by stress such as an antibiotic. Because of the inherent stochasticity, the probability of a population containing zero persisters is always positive and thus the risk of extinction is always present. From here on we will refer to $\sum_g P_{g,0}(t)$ as the population risk, as it is the total probability of zero persisters at time t. There are two main questions that arise concerning the risk: First, given initial conditions at $t = 0$, what is the risk at any $t > 0$? Second, given a population of any size of which we do not know the history, what is the risk? The first question corresponds to growing a culture in a laboratory, in which we know the exact initial conditions. The second corresponds to populations that we encounter outside of the lab, for example inside the human body, in which we do not know the history but only the current population size. Both answers of course depend on the reaction parameters and the fraction of persister cells. In case both these questions cannot be answered, a third question arises: What parameters influence the risk and in what way? We will focus on this question later, and try to answer the two main questions first.

These first two questions are of course not independent of each other. Given the risk at a certain population size, we can find the risk at time t if we know the distribution of total population size at this time. Unfortunately, finding exact probability distributions, i.e. solving the master equation, is impossible for all but the simplest of models [23]. Because of this, we will have to find other ways of studying the persister cell model. A comparable process is that of exponential growth. As persister cell fractions are often very small, the dynamics of growing cells are closely related to those of exponential growth. Thus, we would like to be able to find the exact probability distribution of exponential growth, which is a negative binomial distribution as we derive in Section 2. Using this, we can find an exact expression for the probability generating function of a simplified model of persisters, which we do in Section 3.1. We show numerically that this simplified model is an accurate approximation of the true persister model in Section 3.2.

Since we know the exact distribution of exponential growth, we expect the distribution of the number of growing cells to be similar to that. In Section 5 we use the Gillespie algorithm [18] to simulate the persister cell process. We then fit distributions to the obtained data to show that the distribution of the number of growing cells is in fact similar to that of exponential growth. We also show that the distribution of persister cells at time t fits a negative binomial distribution, and we show the distribution of persister cells given a fixed

total number of cells in Section 5.3. However, since we cannot find exact solutions to the full probability distributions we have to study other properties of the model. Thus we now have to find answers to our third question.

Two properties clearly related to the risk of extinction are the growth rate of the system and the fraction of persister cells $\phi = \frac{p}{g+p}$, which is equivalent to the ratio $\gamma = \frac{g}{p}$ by the relation $\phi = \frac{1}{1+\gamma}$. Another property related to the risk is the noise $\eta = \frac{Var(p)}{\mathbb{E}[p]^2}$ in the number of persister cells. Noise is interesting as it is directly related to the risk, as higher noise means higher risk. The fraction, growth rate and noise are all dependent on each other: Naturally, lower persister fractions lead to higher growth rate but higher risk. The fraction appears to converge towards a steady state as we show in Section 4.1. In Section 4.1.2 we show the interesting result that the noise in the number of persister cells equals the noise in the number of growing cells. The noise appears to become constant when the fraction becomes constant, as we show in Section 4.1.3. However, the steady state value for the noise depends not only on the reaction rates but very heavily on the initial conditions. This is shown numerically in Section 4.2 and derived exactly for exponential growth in Section 2. Using the derived expression for the fraction, we derive an exact probability distribution for a process in which the fraction is constant in Section 4.3.

An overview of all results is given in the table below.

Result	Section
Noise in exponential growth depends on initial conditions	2.1
Exact probability distribution of exponential growth	2.2
Noise in exponential growth as derived from the master equation	2.3
Probability generating function for the simplified model	3
Risk for the simplified model	3.1.4
Fitting of distributions to numerically obtained data	5
Derivation of the steady-state fraction	4.1.1
Noise in persister cells equals noise in growing cells	4.1.3
Noise converges towards a steady-state	4.1.3
Visualisation of dependence of noise on initial conditions	4.2
Distribution in constant fraction	4
Risk in constant fraction	4.3

2. Exponential growth

A simple process similar to persister cell populations is the process of exponential growth. In fact, if we set $v_1 = v_2 = 0$ in (1.1) and assume $p = 0$ at time 0, we get the process of only growing cells. Since the fraction of persister cells is usually very low ($10^{-2} - 10^{-6}$), the population of growing cells in the persister model behaves very similar to exponential growth. Whereas for the persister process finding an exact solution is not always feasible, for exponential growth we are able to find the exact probability distribution at any time t given the initial conditions.

2.1. Noise in indefinitely growing systems

One interesting aspect of exponential growth is the dependence of the noise on the initial conditions. In most biological systems, the system will converge towards a steady state and the noise in that state depends only on the state itself, not on the initial conditions. However, during exponential growth there is no steady state as the number of particles continuous to grow indefinitely. This causes the noise to be heavily dependent on the conditions at $t = 0$, as we will show now.

Theorem 2.1. *In an exponential growth process with initial number of cells N_0, the noise at any time t is equal to*

$$\frac{1}{N_0}\eta_{n_t}.$$

where n_t is an exponential growth process with 1 cell at time 0.

Proof. Suppose that we have an exponential growth process that starts with a single cell at $t = 0$. Naturally, this process shows stochasticity as the rate of division is a random variable. At some time t the number of cells n_t has some distribution with expectation $\langle n_t \rangle$ and some variance σ_t^2. Values for $\langle n_t \rangle$ and σ_t^2 are not strictly necessary for our purpose, but we will derive them later for exponential growth.

Now consider a process N_t that starts with N_0 cells. Since all cells grow independent of each other, this process is equal to the sum of N_0 processes that start with a single cell. Since these N_0 are independent and identically distributed, their expectation is

$$\mathbb{E}[N_t] = \mathbb{E}[\sum_{i=1}^{N_0} n_t] = \sum_{i=1}^{N_0} \mathbb{E}[n_t] = \sum_{i=1}^{N_0} \langle n_t \rangle = N_0 \langle n_t \rangle.$$

And the variance of the sum is the sum of the variances:

$$Var(N_t) = Var(\sum_{i=1}^{N_0} n_t) = \sum_{i=1}^{N_0} Var(n_t) = N_0\sigma_t^2.$$

So that the noise is

$$\eta = \frac{Var(N_t)}{\mathbb{E}[N_t]^2} = \frac{N_0\sigma_t^2}{(N_0\langle n_t\rangle)^2} = \frac{1}{N_0}\frac{\sigma_t^2}{(\langle n_t\rangle)^2} = \frac{1}{N_0}\eta_{n_t}.$$

□

Since this is true for the exponential growth process, surely it must hold for the persister model as well! In fact, a much more general result here is that for any system in which cells grow independent of each other, the noise depends on the initial conditions. In the next section we will show the exact distribution of exponential growth to further exemplify this.

2.2. Probability distribution of exponential growth

Consider an exponential growth process that starts with a single cell $n_0 = 1$ and has growth rate μ, so that the time in which each cell divides is exponentially distributed with parameter μ and mean $\frac{1}{\mu}$. We are interested in finding $P(n_t = k)$ for all $t > 0$. An exact solution to this was first found by the physicist W.H. Furry in 1937 [20] as a solution to high-energy electrons that multiply when passing through a sheet of lead. Furry solved the master equation related to this process by direct integration and induction. We derive the master equation and show Furry's method in the next section. Another solution was found by D.G. Kendall in 1949 [22], who used the master equation to derive and solve a partial differential equation for the generating function as we show in Section A.5. In this section we show that it is possible to find a solution without the master equation. Note that there are several parametrizations for the geometric and negative binomial distributions, which is important to consider when using them. See also Section A.4

Theorem 2.2. *Let N_t be an exponential growth process with initial value N_0 and growth rate μ. Then at time t, the number of cells is given by a negative binomial distribution with parameters N_0 and $e^{-\mu t}$:*

$$P(N_t = k) = \binom{k-1}{N_0-1}(e^{-\mu t})^{N_0}(1-e^{-\mu t})^{k-N(0)}.$$

To prove this theorem, we first need to find the distribution for an exponential growth process that starts with a single cell. The exact source of this proof is unknown to us, but can be found online [21].

Lemma 2.3. *Let n_t be an exponential growth process with $n_0 = 1$ and growth rate μ. Then the distribution of the number of cells at any time t is given by a geometric distribution with parameter $e^{-\mu t}$:*

$$P(n_t = k) = (1 - e^{-\mu t})^k e^{-\mu t}. \tag{2.1}$$

Proof. Let T_k be the time at which the k'th cell is created, and set $T_1 = 0$. We see that $P(n_t \geq k) = P(T_k \leq t)$. Now we look at the same problem from a different angle: if at time t we have k cells and we go backwards in time, what is the probability that $k-1$ of these cells where created in the interval $(0, t]$? It is fairly easy to see that this probability equals $P(T_k \leq t)$. But this process is equivalent to a decay process that starts with $k-1$ cells at time 0, all of which decay with rate μ. Then $P(T_k \leq t)$ is equal to the probability of the equivalent process being extinct at time t. Since for each single cell the probability of surviving until time t is $e^{-\mu t}$, the probability of having 0 cells at time t equals $(1 - e^{-\mu t})^{k-1}$, which is the maximum of $k-1$ exponentially distributed variables. Thus $P(n_t \geq k) = (1 - e^{-\mu t})^{k-1}$, meaning $P(n_t \leq k) = 1 - (1 - e^{-\mu t})^k$. It is well known that this cumulative distribution function corresponds to a probability mass function $P(n_t = k) = (1 - e^{-\mu t})^k e^{-\mu t}$ which is a geometric distribution with parameter $e^{-\mu t}$, corresponding to (2.1) □

With this, we can prove Theorem 2.2:

Proof (of Theorem 2.2). If instead of starting with a single cell we start with N_0 cells, we can consider this as N_0 independent geometrically distributed processes by the above lemma. Let $X_1, X_2, \ldots, X_{N_0}$ be i.i.d. geometrically distributed variables with parameter $q = e^{-\mu t}$. We claim that their sum $S_{N_0}(t)$ has negative binomial distribution with probability mass function

$$P(S_{N_0} = k) = \binom{k-1}{N_0 - 1} q(t)^{N_0} (1 - q(t))^{k - N(0)}.$$

For $N_0 = 1$ this is trivial. By induction:

$$\begin{aligned}
P(S_{N_0+1} = k) &= \sum_{n=N_0}^{k-1} P(S_{N_0} = n \cap X_{N_0+1} = k - n) = \sum_{n=N_0}^{k-1} P(S_{N_0} = n) P(X_{N_0+1} = k - n) \\
&= \sum_{n=N_0}^{k-1} \binom{n-1}{N_0 - 1} q^{N_0} (1-q)^{n - N_0} (1-q)^{k-n-1} q \\
&= q^{N_0+1} (1-q)^{k-(N_0+1)} \sum_{n=N_0}^{k-1} \binom{n-1}{N_0 - 1} \\
&= \binom{k-1}{N_0} q^{N_0+1} (1-q)^{k-(N_0+1)}.
\end{aligned} \tag{2.2}$$

Note that for $N_0+1 \geq k$ this still holds, as the sum will be empty. The last equality might not be trivial, so we show by induction on k that $\sum_{n=N_0}^{k-1} \binom{n-1}{N_0-1} = \binom{k-1}{N_0}$. For $k = 1$ it is trivial. For $k > 1$:

$$\begin{aligned}\sum_{n=N_0}^{k-1} \binom{n-1}{N_0-1} &= \binom{k-2}{N_0-1} + \sum_{n=N_0}^{k-2} \binom{n-1}{N_0-1} \\ &= \binom{k-2}{N_0-1} + \binom{k-2}{N_0} = \binom{k-1}{N_0}.\end{aligned}$$

□

Continuing the result from the previous section, the mean and variance of the negative binomial (N_0, q) distribution are $\frac{N_0}{q}$ and $\frac{N_0(1-q)}{q^2}$, thus the noise is $\eta = \frac{(1-q)}{N_0}$. For large t, $q \to 0$ so $\eta \to \frac{1}{N_0}$, again showing that the noise depends on the initial conditions.

In Section 3.1 we use the probability generating function of exponential growth, which is

$$F_e(s,t) = \mathbb{E}[s^{S_{N_0}(t)}] = \frac{(e^{-\mu t}s)^{N_0}}{(1-(1-e^{-\mu t})s)^{N_0}}. \tag{2.3}$$

Naturally, this is also the product of N_0 times the pgf of the geometric distribution with parameter $e^{-\mu t}$, as the pgf of a sum of independent variables is simply the product of the individuals pgf's.

2.3. The master equation of exponential growth

A third way of studying exponential growth is by the master equation. Although it may seem redundant to do so as we already found the exact probability distribution, in more complex systems the probability distribution is often unknown. The master equation is always is always known, so the method of studying moments using the master equation as we will show here is always viable.

For a system of only growing cells, we have the master equation:

$$\frac{dP_n}{dt} = \mu(n-1)P_{n-1} - \mu n P_n. \tag{2.4}$$

By direct integration and induction this equation can be solved [20].

Alternative proof of lemma 2.3. Assuming $P_1(0) = 1$ we have

$$\frac{dP_1}{dt} = -\mu P_1.$$

which has solution $P_1(t) = e^{-\mu t}$. We claim that for $n > 1$ we have $P_n = (1-e^{-\mu t})^{n-1}e^{-\mu t}$. For $n = 1$ we just showed that this holds. By induction:

$$\frac{dP_n}{dt} = \mu(n-1)P_{n-1} - \mu n P_n = \mu(n-1)(1-e^{-\mu t})^{n-2}e^{-\mu t} - \mu n P_n.$$

This is a linear ordinary differential equation, of which the solution can be found in any introductory ODE or Calculus book. The solution is

$$\begin{aligned}
P_n(t) &= e^{-n\mu t} \int_{s=0}^{t} \mu(n-1)e^{-\mu s}(1-e^{-\mu s})^{n-2}e^{n\mu s}ds \\
&= e^{-n\mu t} \int_{s=0}^{t} \mu(n-1)e^{-(n-1)\mu s}(1-e^{-\mu s})^{n-2}ds \\
&= e^{-n\mu t} \int_{s=0}^{t} \mu(n-1)(e^{\mu s}-1)^{n-2}e^{\mu s}ds \\
&= e^{-n\mu t}|(e^{\mu s}-1)^{n-1}|_{s=0}^{s=t} \\
&= e^{-n\mu t}(e^{\mu t}-1)^{n-1} \\
&= e^{-\mu t}(1-e^{-\mu t})^{n-1},
\end{aligned}$$

which is the geometric distribution as we showed before. □

Next to the result above, we can use the master equation (2.4) to find the time derivatives of the first and second moments. With this we can show that the noise depends on the initial conditions.

Alternative proof of Theorem 2.1.

$$\begin{aligned}
\frac{d\langle n\rangle}{dt} &= \sum_{n=0}^{\infty} n\frac{dP_n}{dt} = \sum_{n=0}^{\infty} n(\mu(n-1)P_{n-1} - \mu n P_n) \\
&= \mu\sum_{n=0}^{\infty} n(n-1)P_{n-1} - n^2P_n = \mu\sum_{n=0}^{\infty}((n+1)n-n^2)P_n \\
&= \mu\sum_{n=0}^{\infty} nP_n = \mu\langle n\rangle.
\end{aligned}$$

and

$$\begin{aligned}
\frac{d\langle n^2\rangle}{dt} &= \sum_{n=0}^{\infty} n^2\frac{dP_n}{dt} = \sum_{n=0}^{\infty} n^2(\mu(n-1)P_{n-1} - \mu n P_n) \\
&= \mu\sum_{n=0}^{\infty} n^2(n-1)P_{n-1} - n^3P_n = \mu\sum_{n=0}^{\infty}((n+1)^2n-n^3)P_n \\
&= \mu\sum_{n=0}^{\infty}(2n^2+n)P_n = 2\mu\langle n^2\rangle + \mu\langle n\rangle.
\end{aligned}$$

Since this system of moments is closed, we can write it in matrix form:

$$\frac{d}{dt}\begin{pmatrix}\langle n\rangle\\ \langle n^2\rangle\end{pmatrix} = \begin{pmatrix}\mu & 0\\ \mu & 2\mu\end{pmatrix}\begin{pmatrix}\langle n\rangle\\ \langle n^2\rangle\end{pmatrix}.$$

The solutions to this system are $\begin{pmatrix}\langle n\rangle\\ \langle n^2\rangle\end{pmatrix} = c_1 v_1 e^{\lambda_1 t} + c_2 v_2 e^{\lambda_2 t}$ for λ_i and v_i respectively the eigenvalues and eigenvectors of the matrix, and c_i constants to be determined by the initial values. We assume that at $t = 0$ we have a known number of cells N_0, thus $\langle n_0\rangle = N_0$ and $\langle n_0^2\rangle = N_0^2$. The eigenvalues of the matrix are μ and 2μ with corresponding eigenvectors $\begin{pmatrix}1\\ -1\end{pmatrix}$ and $\begin{pmatrix}0\\ 1\end{pmatrix}$. Thus we have at $t = 0$:

$$\begin{pmatrix}N_0\\ N_0^2\end{pmatrix} = c_1\begin{pmatrix}1\\ -1\end{pmatrix} + c_2\begin{pmatrix}0\\ 1\end{pmatrix},$$

which has solution $c_1 = N_0$ and $c_2 = N_0^2 + N_0$. So for all t

$$\begin{pmatrix}\langle n\rangle\\ \langle n^2\rangle\end{pmatrix} = \begin{pmatrix}N_0 e^{\mu t}\\ (N_0^2 + N_0)e^{2\mu t} - N_0 e^{\mu t}\end{pmatrix}.$$

In the limit $t \to \infty$ the negative term vanishes as its exponent is lower. Thus for large t the noise becomes

$$\eta = \frac{\langle n^2\rangle - \langle n\rangle^2}{\langle n\rangle^2} = \frac{(N_0^2 + N_0)e^{2\mu t} - N_0 e^{\mu t} - N_0^2 e^{2\mu t}}{N_0^2 e^{2\mu t}} = \frac{N_0 e^{2\mu t} - N_0 e^{\mu t}}{N_0^2 e^{2\mu t}} \to \frac{1}{N_0},$$

showing again that the noise is dependent on the initial conditions. □

A comparison of the noise shows that it is the same as in the previous section:

$$\frac{1-q}{N_0} = \frac{1 - e^{-\mu t}}{N_0} = \frac{N_0 e^{2\mu t} - N_0 e^{\mu t}}{N_0^2 e^{2\mu t}}.$$

However, as said before, in more complex systems the exact probability distribution is often unknown whereas the master equation is known. Thus the moment equation method is always viable.

3. Extinction risk in a simplified model

In this section we derive an exact solution for the probability generating function of a simplified model for persister cells. Although this model is simpler than the 'true' model, the difference between the models is extremely small. Because of this, the simplified model will still result in an accurate caculation of the risk. The accuracy of the simplified model depends on the fraction of persister cells. Since persister cell fractions are often very small, we define the simplified model by a small adjustment to the true model: If a switch from growing to persisting cells does not change the number of growing cells, i.e. a switch from growing to persisting magically creates a persister cell and the reverse switch only kills a persister without creating a growing cell, then we have a simpler model. The accuracy of this model will be shown numerically in Section 3.2. In Section 3.1, we translate it to a branching process to derive an exact solution for its generating function. The main result of this section is the following theorem:

Theorem 3.1. *Let $P^*_{g,p}$ be the probability distribution of the simplified persister model. The generating function of the process with initial conditions $g = 1$, $p = 0$ is*

$$F_g(s_1, s_2, t) = \sum_{g,p \in \mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}} P^*_{g,p}(t) s_1^g s_2^p.$$

and satisfies

$$F_g(s_1, s_2, t) = \frac{e^{-\mu_g t + \frac{v_1(s_2-1)}{v_2}(1-e^{-v_2 t})}}{\frac{1}{s_1} + \frac{\mu_g}{v_2} e^{\frac{v_1(s_2-1)}{v_2}} \frac{1}{(\frac{-v_1(s_2-1)}{v_2})^{\frac{\mu_g}{v_2}}} \int_{\frac{v_1(s_2-1)}{v_2}}^{\frac{v_1(s_2-1)}{v_2} e^{-v_2 t}} x^{\frac{\mu_g}{v_2}-1} e^{-x} dx}. \tag{3.1}$$

3.1. Exact solution of the generating function

3.1.1. Introduction to Branching processes

The best known branching process is the discrete Galton-Watson branching process [32]. This process counts the number of particles or nodes Z_n at time n. Suppose we start at time $n = 0$ with a single cell, $Z_0 = 1$. At time n all Z_n nodes die and each node produces X offspring, where X is a random variable. Thus $Z_{n+1} = \sum_{i=1}^{Z_n} X_i$ with all X_i independent

and identically distributed. As the next timestep depends only on the current number of cells, the Galton-Watson process is a Markov jump process.

There are many possible extensions to the Galton-Watson process. First of all, it is possible for Z_n to be vector-valued, which also causes X to be a multi-dimensional random variable. This allows the branching process to model processes that contain multiple types of cells. Another more significant change is the extension to continuous time branching processes. A lot of processes do not have events at set intervals, but at random times. In the case where these random times are exponential random variables, the branching process is also a Markov process. We can apply results for continuous time branching processes to the persister cell model, as that is also a Markov process. We will not go into the details here but more theory can be found in works such as Kimmel and Axelrod [32] and Harris [33].

3.1.2. Branching process of the persister model

We can translate the models of persister cells into a branching process. Naturally, we have two types of particles: growing (g) type and persister (p) type. Since cells do not die in our model, we do not have a death rate. However, we can consider a cell dividing to be the event of a cell dying and producing 2 cells, and the event of a switch as the death of one type and the birth of the other type. Because the growth and switching rates are exponential, we can sum them to get the branching process parameters. Thus the lifespans of cells are distributed exponentially with means $\lambda_g = v_1 + \mu_g$ and $\lambda_p = v_2 + \mu_p$. For the true model each type g particle produces 2 type g particles with probability $1 - \alpha$ and a type p particle with probability α at death, where $\alpha = \frac{v_1}{\mu_g + v_1}$. Each type p particle produces a type g particle at death (assuming $\mu_p = 0$).

For the simplified model in which switching does not change the number of growing cells, a type g particle produces 2 type g particles with probability $1 - \alpha$ and one type g and one type p particle with probability α. Also, a type p particle produces no offspring in this model.

For each time t, the two processes corresponding to these models are random variables. This means that they have a probability generating function (pgf), defined by

$$F(s_1, s_2, t) = \sum_{g,p=0}^{\infty} P(g \text{ growing, } p \text{ persisters}|t) s_1^g s_2^p.$$

A solution for the pgf is a solution for the full probability distribution. We will find this solution in next section.

3.1.3. Methods for solving the probability generating function

Let $F_i(s_1, s_2, t), i \in \{g, p\}$ be the pgf for the branching process that starts with one type i cell at $t = 0$. Let λ_i be the lifetime of particle i, and let $f_i(s_1, s_2)$ be the joint pgf of

the number of offspring generated by a type i particle. We have a system of ordinary differential equations [32, eq. 4.9]

$$\frac{dF(s,t)}{dt} = -\lambda \cdot [F(s,t) - f(F(s,t))].$$

in which F, f, s and λ are all two-dimensional and $\cdot$ is the inner product between vectors. The initial condition is $F(s,0) = s$. In the true model, a growing cell generates two growing cells with probability $1-\alpha$ and a single persister cell with probability α, thus $f_g(s_1,s_2) = (1-\alpha)s_1^2 + \alpha s_2$. Likewise, $f_p(s_1,s_2) = s_1$ as a persister cell always generates a growing cell. Thus the system of differential equations is

$$\begin{aligned} \frac{dF_g}{dt} &= -\lambda_g F_g + \lambda_g((1-\alpha)F_g^2 + \alpha F_p) \\ \frac{dF_p}{dt} &= -\lambda_p F_p + \lambda_p F_g \end{aligned} . \tag{3.2}$$

Unfortunately, this system of differential equations is not easily solvable, and perhaps even impossible to solve with known methods. For this exact reason we introduced the simplified model. For the simplified model we have $f_g(s_1,s_2) = (1-\alpha)s_1^2 + \alpha s_1 s_2$ and $f_p(s_1,s_2) = 1$. This results in the system of differential equations

$$\begin{aligned} \frac{dF_g}{dt} &= -\lambda_g F_g + \lambda_g((1-\alpha)F_g^2 + \alpha F_p F_g) \\ \frac{dF_p}{dt} &= -\lambda_p F_p + \lambda_p \end{aligned} . \tag{3.3}$$

Here we see why the simplified model is useful: The true model admits a system of differential equations in which both F_g and F_p depend on each other, while in the simplified model F_p is completely independent of F_g! This allows us to first find a solution for F_p and substitute that in the equation for F_g. The resulting equation is a type of Bernoulli-equation, which has been intensively studied and is solvable as we will show below.

3.1.4. Proof of Theorem 3.1

With the system of differential equations from the previous section, we can prove the theorem given in the introduction. More specifically, we show that (3.1) is a solution to (3.3).

Proof of Theorem 3.1. Since F_p is independent of F_g for the simplified model as seen in (3.3), we can find a general solution. Note that F_p, $F_p(t)$ and $F_p(s,t)$ are all used to indicate the same function. We have

$$\frac{dF_p}{dt} = -\lambda_p F_p + \lambda_p,$$

which we rewrite to

$$\frac{dF_p}{F_p - 1} = -\lambda_p dt.$$

Integrating both sides gives

$$\ln(F_p(t)-1)-\ln(F_p(0)-1)=-\lambda_p t.$$

As we start with a single particle, $F_p(0,s_2)=s_2$. Using this and taking the exponential gives us

$$(F_p(t)-1)(s_2-1)^{-1}=e^{-\lambda_p t},$$

which yields the solution

$$F_p(t,s_2)=e^{-\lambda_p t}(s_2-1)+1. \tag{3.4}$$

Plugging in (3.4) in (3.3) yields

$$\frac{dF_g}{dt}=F_g(-\lambda_g+\lambda_g\alpha(e^{-\lambda_p t}(s_2-1)+1))+F_g^2\lambda_g(1-\alpha).$$

Substituting $w(t)=\frac{1}{F_g(t)}$ gives

$$\dot{w}=[\lambda_g-\lambda_g\alpha(e^{-\lambda_p t}(s_2-1)+1))]w-\lambda_g(1-\alpha).$$

Which has solution (see Polyanin and Zaitsev [24])

$$w(t)=C_1e^{-K}-e^{-K}\int e^K\lambda_g(1-\alpha)dt.$$

Here $K(t)=-\int\lambda_g-\lambda_g\alpha(e^{-\lambda_p t'}(s_2-1)+1)dt'$. We solve this integral:

$$\begin{aligned}K(t)&=-\int\lambda_g-\lambda_g\alpha(e^{-\lambda_p t'}(s_2-1)+1)dt'\\&=-\lambda_g t(1-\alpha)+\frac{\lambda_g\alpha(s_2-1)}{\lambda_p}(1-e^{-\lambda_p t}).\end{aligned}$$

Now to find F_g we need to calculate

$$\begin{aligned}\int\exp(e^{K(t)})dt&=\int\exp(e^{-\lambda_g t(1-\alpha)+\frac{\lambda_g\alpha(s_2-1)}{\lambda_p}(1-e^{-\lambda_p t})})dt\\&=e^{\frac{\lambda_g\alpha(s_2-1)}{\lambda_p}}\int e^{-\lambda_g t(1-\alpha)}e^{-\frac{\lambda_g\alpha(s_2-1)}{\lambda_p}e^{-\lambda_p t}}dt.\end{aligned}$$

Let $a=-\lambda_g(1-\alpha)$, $b=\frac{-\lambda_g\alpha(s_2-1)}{\lambda_p}$ and $c=-\lambda_p$ so that we need to solve $\int e^{at+be^{ct}}$. First, we substitute $u=e^t$ to get

$$\int e^{at+be^{ct}}dt=\int\frac{1}{u}e^{a\ln(u)+bu^c}du=\int u^{a-1}e^{bu^c}du.$$

Now we substitute $v=b^{\frac{a}{c}}u^a$ to get

$$\int u^{a-1}e^{bu^c}du=\frac{1}{ab^{\frac{a}{c}}}\int e^{v^{\frac{c}{a}}}dv.$$

This is a standard integral, which is given by

$$\int e^{v^{\frac{c}{a}}} dv = \frac{a}{c}\Gamma(\frac{a}{c}, -v^{\frac{c}{a}}).$$

Here Γ is the incomplete gamma function defined by $\Gamma(s,x) = \int_x^\infty t^{s-1}e^{-t}dt$. Note that we need to integrate from 0 to t. Since $v = -be^{ct}$ we get

$$\int_{-b}^{-be^{ct}} e^{v^{\frac{c}{a}}} dv = \frac{a}{c}(\Gamma(\frac{a}{c}, -be^{ct}) - \Gamma(\frac{a}{c}, -b)) = \frac{a}{c}\int_{-b}^{-be^{ct}} x^{\frac{a}{c}-1}e^{-x}dx.$$

Substituting everything back we get

$$\begin{aligned} w(t) &= e^{-K}(C_1 - \lambda_g(1-\alpha)\int e^K dt) \\ &= \frac{C_1 + \lambda_g(1-\alpha)e^{\frac{\lambda_g\alpha(s_2-1)}{\lambda_p}} \frac{1}{\lambda_p(\frac{-\lambda_g\alpha(s_2-1)}{\lambda_p})^{\frac{\lambda_g(1-\alpha)}{\lambda_p}}} \int_{\frac{\lambda_g\alpha(s_2-1)}{\lambda_p}}^{\frac{\lambda_g\alpha(s_2-1)}{\lambda_p}e^{-\lambda_p t}} x^{\frac{\lambda_g(1-\alpha)}{\lambda_p}-1}e^{-x}dx}{e^{-\lambda_g t(1-\alpha)+\frac{\lambda_g\alpha(s_2-1)}{\lambda_p}(1-e^{-\lambda_p t})}}. \end{aligned}$$

Since $F_g(0) = s_1$ and thus $w(0) = \frac{1}{s_1}$, we must have $C_1 = \frac{1}{s_1}$. Now substituting the original reaction parameters back we get (3.1). □

We should have $F_g(t,1,1) = 1$ regardless of t, as it is simply the sum of all probabilities. However, if $s_2 = 1$ it is unclear what exactly happens. However, if $\alpha = 0$ in (3.1), this should give the pgf of the geometric distribution (exponential growth), which is given by $F_e(s_1) = \frac{s_1 e^{-\lambda_g t}}{1-s_1(1-e^{-\lambda_g t})}$. With regards to the second term in the denominator, setting $\alpha = 0$ should make it behave similarly to setting $s_2 = 1$, as they both set the same terms to zero. Normally, calculating this limit would be hard but because we know what happens if $\alpha = 0$, we can find an expression for $s_2 = 1$. So if $s_2 = 1$ we have

$$F_g(t, s_1, 1) = \frac{s_1 e^{-\lambda_g t}}{1 - s_1(1 - e^{-\lambda_g t})}.$$

And setting $s_1 = 1$ in this indeed results in $F_g(t,1,1) = 1$. By Theorem 3.1 we can find the risk for this model:

Corollary 3.2. *The probability of zero persisters cells is given by $F_g(t, s_1 = 1, s_2 = 0)$. This is*

$$\sum_{g=0}^{\infty} P_{g,0}(t) = F_g(t,1,0) = \frac{e^{-\mu_g t - \frac{v_1}{v_2}(1-e^{-v_2 t})}}{1 + \frac{\mu_g}{v_2}e^{\frac{v_1}{v_2}} \frac{1}{(\frac{v_1}{v_2})^{\frac{\mu_g}{v_2}}} \int_{\frac{v_1}{v_2}e^{-v_2 t}}^{\frac{v_1}{v_2}} x^{\frac{\mu_g}{v_2}-1}e^{-x}dx}. \tag{3.5}$$

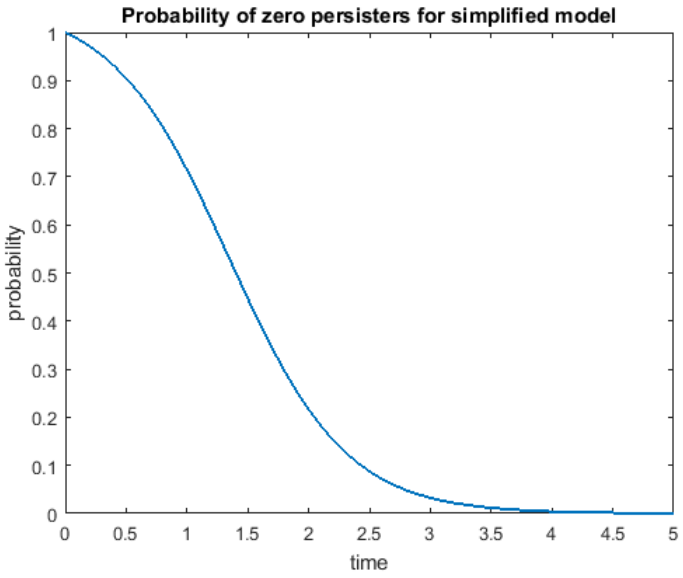


Figure 3.1.: Probability of zero persister cells for the simplified model, as calculated with (3.5). In this figure $\mu_g = 2$, $\mu_p = 0$, $v_1 = 0.12$ and $v_2 = 0.1$.

A plot of this is shown in Figure 3.1 for parameter values $\mu_g = 2h^{-1}$, $v_1 = 0.12h^{-1}$, $v_2 = 0.1h^{-1}$ and $\mu_p = 0h^{-1}$ as in [2]. However, we changed v_1 by a factor 10^5 to be able to numerically compare this to the true model in the next section.

Now that we know the risk from (3.5), we are interested in how the reaction parameters influence the risk. Clearly, increasing μ_g reduces the risk. This should be trivial, as higher growth rate causes more growing cells, which create more persister cells. Unfortunately, cells cannot simply increase their growth rate. Thus, we would like to know how v_1 and v_2 influence the risk, as cells can control these rates. To filter out the effect of the growth rate, we rescale by it. This means that we take $t^* = \mu_g t$, $v_1^* = \frac{v_1}{\mu_g}$ and $v_2^* = \frac{v_2}{\mu_g}$. If we then leave out the * for clarity, we get

$$P(\text{zero persisters}|t) = \frac{e^{-t-\frac{v_1}{v_2}(1-e^{-v_2 t})}}{1 + \frac{1}{v_2} e^{\frac{v_1}{v_2}} \frac{1}{(\frac{v_1}{v_2})^{\frac{1}{v_2}}} \int_{\frac{v_1}{v_2} e^{-v_2 t}}^{\frac{v_1}{v_2}} x^{\frac{1}{v_2}-1} e^{-x} dx}. \tag{3.6}$$

From this is it still unclear how v_1 and v_2 affect the risk. For this reason we will plot the risk for different values of these parameters, taking t fixed at 1 (so $t^* = 2$). In Figure 3.2 we see that an increase or decrease in v_1 corresponds to a decrease or increase in the risk respectively. This is as expected, as higher v_1 means that the probability of a persister cell being created is higher. The rate v_2 does not seem to influence the risk very much, especially when either v_1 or v_2 is low. Both of these have an intuitive explanation: If v_1 is low, almost no persister cells are created so there are simply no persister cells that can

switch back, rendering the impact of v_2 minimal. This is also seen by the fact that the risk is almost 1 for low v_1. In the case where v_2 is low, i.e. lower than 10^{-1}, switching from persister to growing cells simply almost never occurs. This is because the number of persister cells is already low, so having a low switching rate makes the reverse switch very rare, and thus further decreasing v_2 has little impact.

For v_1 and v_2 both around 0.1-1 however, we see that v_2 does impact the risk significantly. This is better shown in Figure 3.3. We see that if $1 \geq v_1 \geq 0.1$, then changing v_2 between 1 and 0.1 can change the risk by almost 5%. Thus v_2 does have a role in reducing or increasing the risk. It should be noted that the ranges of v_1 and v_2 that impact the risk depend on μ_g and t, which we took fixed here. However, for different values of μ_g and t the plots will look the same, they will just be shifted.

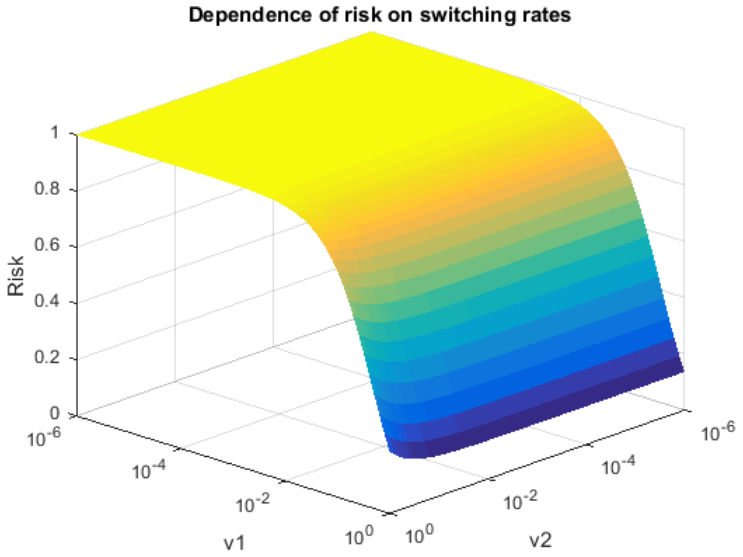


Figure 3.2.: Risk for $\mu_g = 2$ and $t = 1$ with different v_1 and v_2. Clearly, v_2 has very little influence on the risk whereas an increase in v_1 reduces the risk.

3.2. Accuracy of the simplified model

Instead of trying to solve the master equation analytically, we can construct a system of differential equations that represents the master equation. This would be a (double) infinite system which is not solvable, so we set a maximum number of growing and persister cells to make it finite. The finite system is linear, which makes it easy to integrate numerically. However, the finite system only accurately represents the true model if the probability of being in the state with the maximum number of cells is zero, on which we elaborate later in this section.

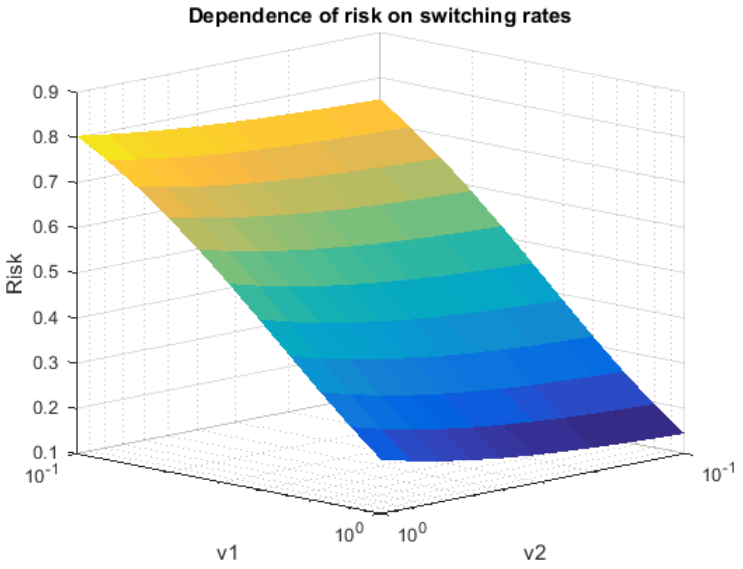


Figure 3.3.: Risk for $\mu_g = 2$ and $t = 1$ with different v_1 and v_2. This is a zoomed in version of Figure 3.2. This figure shows that for $v_1, v_2 \geq 0.1$ the impact of v_2 on the risk is notable.

3.2.1. Mathematical basis for comparison the models

We limit the number of growing and persister cells to constants $gmax$ and $pmax$ respectively. Then we define the vector

$$u = \begin{pmatrix} u_{00} \\ u_{01} \\ \vdots \\ u_{0pmax} \\ u_{10} \\ \vdots \\ u_{gmax,pmax} \end{pmatrix}.$$

such that $u_{gp} = P(\text{g growing and p persister cells})$. The variable u Satisfies the differential equation $\dot{u} = Au$, where A is a matrix defined by the master equation (1.2):

$$\begin{aligned} \frac{\partial u_{g,p}}{\partial t} &= \mu_g((g-1)u_{g-1,p} - gu_{g,p}) + \mu_p((p-1)u_{g,p-1} - pu_{g,p}) \\ &\quad + v_1((g+1)u_{g+1,p-1} - gu_{g,p}) + v_2((p+1)u_{g-1,p+1} - pu_{g,p}). \end{aligned}$$

Since every value of u_{gp} is only dependent on at most 4 other entries of u, the matrix A is sparse. This makes solving the system of differential equations computationally feasible.

Note that for this model we assume $u_{gp} = 0$ if $g > gmax$ or $p > pmax$, or if either $g < 0$ or $p < 0$, which allows us to properly construct A. We also take all values of g and p modulo $gmax$ and $pmax$ respectively, which ensures $\sum_{i,j=0}^{gmax,pmax} u_{ij} = 1$, as there is no flow out of the system to states that we set to 0. For example, once we hit $gmax$ growing cells this modulo turns off the growth rate. This is merely for aesthetics though, as it does not influence the system before the probability of having $gmax$ growing or $pmax$ persister cells is large. However, if we would want to change the model to a physically accurate model of having a limited number of cells (like in an enclosed volume), this change would be necessary, thus for possible future purposes we included it. If, for example, we set $gmax = 3$ and $pmax = 1$ the system is defined by the 8×8 matrix (assuming $\mu_p = 0$):

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -v_2 & v_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & v_2 & -(\mu_g + v_1) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -(\mu_g + v_2) & 2v_1 & 0 & 0 & 0 \\ 0 & 0 & \mu_g & v_2 & -(2\mu_g + 2v_1) & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu_g & 0 & -(2\mu_g + v_2) & 3v_1 & 0 \\ 0 & 0 & 0 & 0 & 2\mu_g & v_2 & -3v_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\mu_g & 0 & 0 \end{pmatrix}. \quad (3.7)$$

Note that the leftmost and rightmost columns are empty, as it is impossible to leave the state with 0 cells and the state with the maximum number of both persisters and growing cells. Also note that the topmost row is empty, as it is impossible to go to the state with zero cells from any other state, as we did not include cell death in our model. Lastly, the sum in each column adds up to zero. This is because it is simply the negative rate of leaving a certain state plus the positive rate of entering all other states from that state.

There is an interesting variation (or simplification) to the above model, as we already showed in Section 3.1. In this variation, which we called the 'simplified' model, the event of a cell switching from the growing to persisting state or vice versa does not change the number of growing cells. However, we would like to study whether it is justified. Thus, we construct the same model as above but now A is defined by a different master equation:

$$\begin{aligned} \frac{\partial u_{g,p}}{\partial t} &= \mu_g((g-1)u_{g-1,p} - gu_{g,p}) + \mu_p((p-1)u_{g,p-1} - pu_{g,p}) \\ &\quad + v_1(gu_{g,p-1} - gu_{g,p}) + v_2((p+1)u_{g,p+1} - pu_{g,p}) \end{aligned}$$

For this model, the matrix from (3.7) changes to

$$A = \begin{pmatrix} 0 & v_2 & v_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -v_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -(\mu_g + v_1) & v_2 & 2v_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -(\mu_g + v_2) & 0 & 2v_1 & 0 & 0 \\ 0 & 0 & \mu_g & 0 & -(2\mu_g + 2v_1) & v_2 & 3v_1 & 0 \\ 0 & 0 & 0 & \mu_g & 0 & -(2\mu_g + 2v_1 + v_2) & 0 & 3v_1 \\ 0 & 0 & 0 & 0 & 2\mu_g & 0 & -3v_1 & v_2 \\ 0 & 0 & 0 & 0 & 0 & 2\mu_g & 0 & -(3v_1 + v_2) \end{pmatrix}.$$

The main visible difference between this and (3.7) is that it is not possible to both enter the zero state and exit the maximum state. The second of these is of no impact, as the system is not accurate when the probability of being in the maximum state differs significantly from zero. The first however, could have some impact, as when the system enters the zero state, it still cannot leave. However, entering the zero state in the simplified model is equal to entering the state $g = 0, p = 1$ in the true model. This probability is extremely small already, and the rate of leaving the state $(0, 1)$ in the true model is v_2. This rate is relatively small as well, thus the state $(0, 1)$ is almost equal to the state $(0, 0)$.

3.2.2. Comparing the simplified and true models

Using the ODE approach shown above we aim to study the differences and similarities between the simplified and true models. In the following figures we used the parameters $\mu_g = 2\ h^{-1}$, $\mu_p = 0\ h^{-1}$, $v_1 = 0.12\ h^{-1}$, $v_2 = 0.1\ h^{-1}$, $pmax = 40$ cells and $gmax = 400$ cells as experimentally found by Kussel et al. [2]. However, we did change v_1 by a factor 10^{-5} to increase the fraction of persister cells. For these values, we expect the resulting fraction of persister cells to asymptotically converge to $\frac{g}{p} \approx 16.6$, or $\frac{p}{g} \approx 0.06$. This might not be a realistic fraction, as in practice the number of persister cells is lower, but increasing $\frac{g}{p}$ would mean we would have to increase $gmax$ as well, thus greatly increasing the computational intesivity of the model. This is unfortunately not feasible, hence the relatively high expected fraction. However, for lower fractions the simplified model should be a better approximation as the switching occurs less. For this reason, results in this section are still relevant.

From Figure 3.4 we see that for $t > 2$ the probability of the state with $gmax$ growing cells is not small, so results are only accurate for $t < 2$ for both models. We are most interested in the probability of having zero persisters. This probability is shown in Figure 3.5. The difference between the models is extremely small, thus we are inclined to conclude that the simplified model is equally suitable for our purpose as the true model. Comparing this to Figure 3.1 they look very similar. Numerical verification shows that the difference between both figures is of order 10^{-9}, which could be attributed either to the slight probability of being in the maximum state in the ODE model, or numerical inaccuracies.

However, there is a difference between the models. In Figure 3.6 we see that the expected number of growing cells does differ significantly. This is as expected, as in the simplified

model the growing cells are independent of the persister cells, so they will simply follow an exponential growth process. The difference in the number of growing cells does change the total number of persister cells and also the fraction of persister cells. This can clearly be seen in figures 3.6 and 3.7.

Although there are differences between the simplified and true model, Figure 3.5 shows that with regards to our research question we can consider the simplified model to be an accurate representation. From this, we can conclude that (3.5) can be used to closely approximate the risk in the persister cell model. We derived this equation only for the process with initial condition $g = 1$, $p = 0$. For a process that starts with $g > 1$ growing cells, the risk is found simply by taking the g'th power of the right hand side of 4(3.5). Unfortunately, this method does not allow us to calculate the risk for populations with initial conditions $p > 0$.

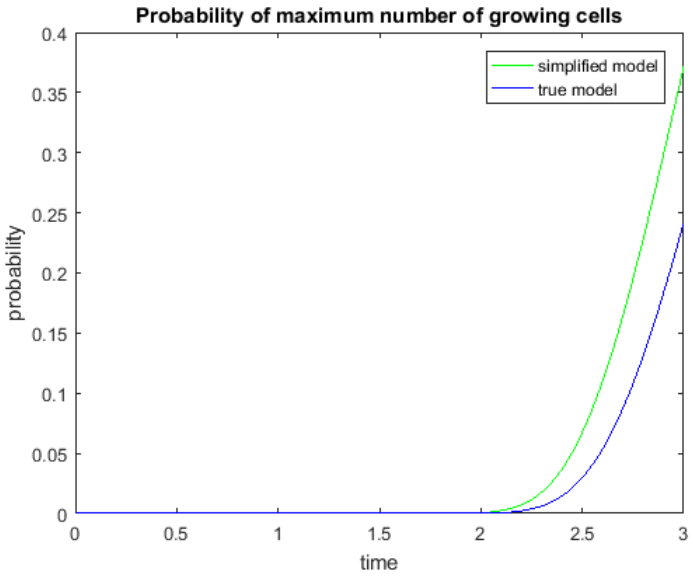


Figure 3.4.: Probabilities of the maximum number of growing cells. Clearly, before time $t < 2$ we can assume both models to be accurate. Since the simplified model grows faster, the probability increases faster as well.

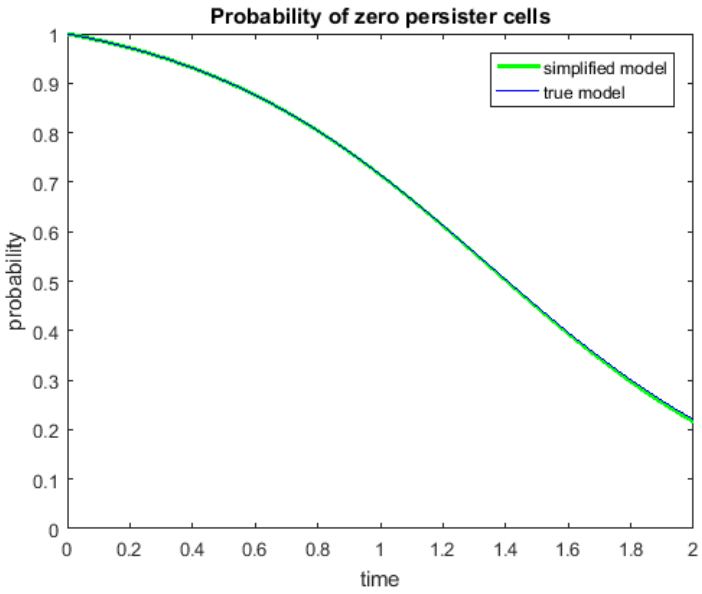


Figure 3.5.: Probability of zero persister cells for both models. The curves are almost equal, showing that the models differ very little in this regard.

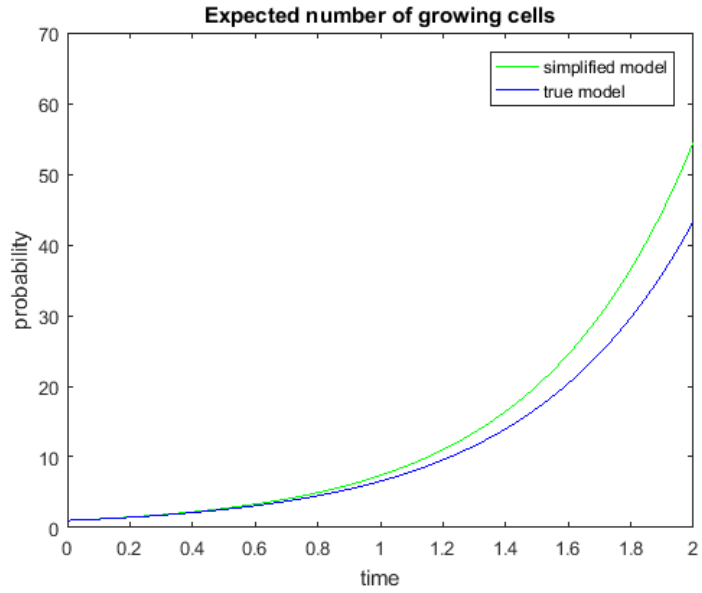


Figure 3.6.: Expected number of growing cells for both models. The simplified models is known to follow an exponential growth curve, while the true model also grows exponentially but with a slightly lower rate.

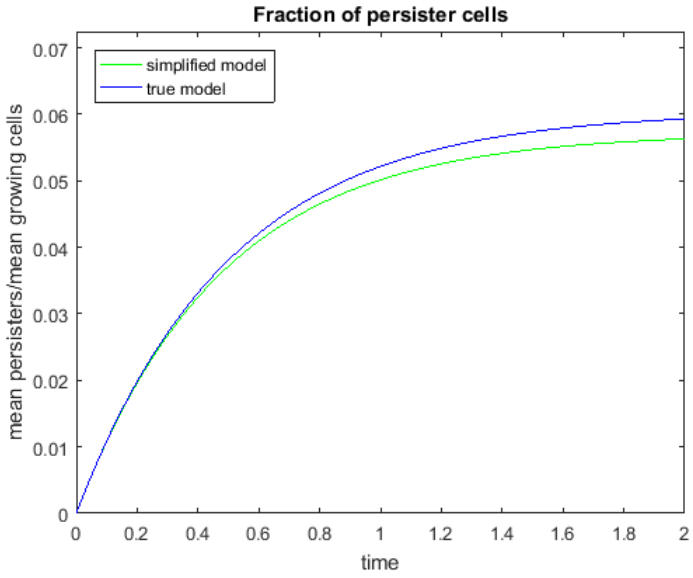


Figure 3.7.: Expected fraction of persister cells for both models. For the simplified model the fraction is slightly lower. This could be explained by the fact that the expected number of growing cells is higher in that model. The expected value towards which the fraction converges is approximately 0.06.

4. Approximation of the risk in a constant fraction model

In Section 2 we derived an exact distribution for the exponential growth process. In this section we again focus on the complete model (1.1) and show that the expectation of the fraction of persister cells converges to a steady state, independent of initial conditions. We also show the exact value of this steady state fraction. Knowing this, we can treat the persister process as an exponential growth process with reduced growth rate, by considering only the total number of cells which does grow exponentially. Using the result from Section 5.3 that persister cells are distributed binomially given the total number of cells, we find an exact solution for the model in which ϕ is constant. In Section 4.1 we show that the noise and fraction converge, and derive the steady state fraction. In Section 4.2 we show that the noise is greatly effected by the initial conditions. Lastly, in Section 4.3 we derive an exact value for the risk in the constant fraction model. The main results of this chapter can be summarized in the following three theorems:

Theorem 4.1. *In the persister model* (1.1) *the expected fraction of persister cells* $\phi = \langle \frac{p}{p+g} \rangle$ *converges in time to a steady state value*

$$\phi = \frac{2v_1}{v_1 + \mu_g - \mu_p + v_2 + \sqrt{(\mu_g - v_1 - \mu_p + v_2)^2 + 4v_1 v_2}}. \tag{4.1}$$

Theorem 4.2. *The expected noise* η_p *in the number of persister cells converges in time to a steady state. Also, the expected noise* η_g *in the number of growing cells converges to that same value.*

Theorem 4.3. *If in the persister model we assume the expectation of the fraction* ϕ *to be constant, then the distribution of the number of cells is*

$$P_{g,p}(t) = \binom{N-1}{N(0)-1} q^{N(0)} (1-q)^{N-N(0)} \binom{N}{p} \phi^p (1-\phi)^{N-p}. \tag{4.2}$$

where $N = g + p$, $q = e^{-\lambda_1 t}$, $\lambda_1 = \frac{1}{2}(\mu_g - v_1 - v_2 + \mu_p + \sqrt{(\mu_g - v_1 - \mu_p + v_2)^2 + 4v_1 v_2})$ *and* ϕ *as in* (4.1)

4.1. Moment master equations

Aside from trying to solve the whole model, it is possible to study the process of persister cells by only using the moments of the probability distribution. As we did in Section 2, we

can use (1.2) to find differential equations for the moments of $P_{g,p}$. Since for every order this yields a closed system of differential equations, it is possible to rewrite it into matrix form and solve the differential equations by calculating the eigenvalues of the matrix. We are interested mostly in the first and second moments, as these are needed to calculate the noise. In this section we also show that the fraction of persister cells will converge to a steady state dependent only on the reaction rates, that the noise will become constant if the fraction does and that the steady state noise in g and p is equal.

4.1.1. First moments

From (1.2) and $\langle p\rangle = \sum_{g,p} gP_{g,p}$, we can find an equation for $\frac{d}{dt}\langle p\rangle$:

$$\begin{aligned}\frac{d}{dt}\sum_{g,p} pP_{g,p} &= \sum_{g,p} p\big(\mu_g[(g-1)P_{g-1,p} - gP_{g,p}] + \mu_p[(p-1)P_{g,p-1} - pP_{g,p}]\\ &\quad + v_1[(g+1)P_{g+1,p-1} - gP_{g,p}] + v_2[(p+1)P_{g-1,p+1} - pP_{g,p}]\big)\\ &= \sum_{g,p} \mu_g[gpP_{g,p} - gpP_{g,p}] + \mu_p[(p+1)pP_{g,p} - p^2P_{g,p}]\\ &\quad + v_1[(p+1)gP_{g,p} - gpP_{g,p}] + v_2[(p-1)pP_{g,p} - p^2P_{g,p}]\\ &= (\mu_p - v_2)\langle p\rangle + v_1\langle g\rangle.\end{aligned} \tag{4.3}$$

Note that this is because $\sum_{g,p}(g-1)P_{g-1,p} = \sum_{g,p} gP_{g,p}$, as the $g=0$ term vanishes, which also occurs if we interchange g and p. Because of symmetry,

$$\frac{d}{dt}\langle g\rangle = (\mu_g - v_1)\langle g\rangle + v_2\langle p\rangle. \tag{4.4}$$

As expected, this gives the closed two-state system

$$\begin{aligned}\frac{d\langle g\rangle}{dt} &= (\mu_g - v_1)\langle g\rangle + v_2\langle p\rangle\\ \frac{d\langle p\rangle}{dt} &= (\mu_p - v_2)\langle p\rangle + v_1\langle g\rangle\end{aligned}. \tag{4.5}$$

Or, written in matrix form,

$$\frac{d}{dt}\begin{pmatrix}\langle g\rangle\\ \langle p\rangle\end{pmatrix} = \underbrace{\begin{pmatrix}\mu_g - v_1 & v_2\\ v_1 & \mu_p - v_2\end{pmatrix}}_{M_1}\begin{pmatrix}\langle g\rangle\\ \langle p\rangle\end{pmatrix}. \tag{4.6}$$

which also could have been derived simply from the reactions in Figure 1.1 as it is the system of differential equations for a deterministic system obeying (1.1). With the solutions to this system, we can derive the steady state fraction.

Proof of Theorem 4.1. Solutions to (4.6) allow us to find an expression for ϕ. Using the notation by Lu et al. [25], who also studied this system, the eigenvalues of M_1 are shown to be $\frac{1}{2}(\sigma_1 + \sigma_2 \pm \Delta)$ with corresponding eigenvectors $\begin{pmatrix} \frac{\sigma_1 - \sigma_2 \pm \Delta}{2v_1} \\ 1 \end{pmatrix}$. Here $\sigma_1 = \mu_g - v_1$, $\sigma_2 = \mu_p - v_2$ and $\Delta = \sqrt{(\sigma_1 - \sigma_2)^2 + 4v_1 v_2}$. For a system of differential equations $\dot{u} = Mu$ the solutions are $\sum_{i=1}^{n} c_i v_i e^{\lambda_i t}$ with n the dimension of u, v_i and λ_i the eigenvectors and eigenvalues of M and c_i constants determined by the initial values. Note that by the Perron-Frobenius Theorem [26] the largest eigenvalue of M_1 has a strictly positive eigenvector.

Let λ_1 be the largest eigenvalue of M_1. Then given initial conditions $g(0)$ and $p(0)$ this yields for our system

$$\begin{pmatrix} \langle g \rangle \\ \langle p \rangle \end{pmatrix} = \begin{pmatrix} c_1 \frac{\sigma_1 - \sigma_2 + \Delta}{2v_1} e^{\lambda_1 t} + c_2 \frac{\sigma_1 - \sigma_2 - \Delta}{2v_1} e^{\lambda_2 t} \\ c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t} \end{pmatrix}, \tag{4.7}$$

with $c_1 = p(0)\frac{\sigma_2 - \sigma_1 + \Delta}{2\Delta} + g(0)\frac{v_1}{\Delta}$ and $c_2 = p(0)\frac{\sigma_1 - \sigma_2 + \Delta}{2\Delta} - g(0)\frac{v_1}{\Delta}$. For large t, $e^{\lambda_1 t} \gg e^{\lambda_2 t}$. As such, we see that

$$\gamma = \frac{\langle g \rangle}{\langle p \rangle} \to \frac{\sigma_1 - \sigma_2 + \Delta}{2v_1}, \text{ or } \phi = \frac{\langle p \rangle}{\langle g \rangle + \langle p \rangle} \to \frac{2v_1}{2v_1 + \sigma_1 - \sigma_2 + \Delta}. \tag{4.8}$$

Thus the ratio and fraction of persister cells depend only on the reaction rates. Substituting back the original parameters yields (4.1). However, we used $\phi = \frac{\langle p \rangle}{\langle g \rangle + \langle p \rangle}$ instead of $\phi = \langle \frac{p}{g+p} \rangle$. We will show in Section 4.1.3 that these converge to the same value. □

4.1.2. Second moments

For the time derivatives of the second moments, $\langle g \rangle$, $\langle p \rangle$ and $\langle gp \rangle$ we can use the same method as for the first moments. This yields

$$\begin{aligned} \frac{d}{dt}\langle g^2 \rangle &= (\mu_g + v_1)\langle g \rangle + v_2 \langle p \rangle + 2(\mu_g - v_1)\langle g^2 \rangle + 2v_2 \langle gp \rangle \\ \frac{d}{dt}\langle gp \rangle &= -v_1 \langle g \rangle - v_2 \langle p \rangle + v_1 \langle g^2 \rangle + (\mu_g + \mu_p - v_1 - v_2)\langle gp \rangle + v_2 \langle p^2 \rangle \\ \frac{d}{dt}\langle p^2 \rangle &= (v_1)\langle g \rangle + (\mu_p + v_2)\langle p \rangle + 2v_1 \langle gp \rangle + 2(\mu_p - v_2)\langle p^2 \rangle. \end{aligned} \tag{4.9}$$

This system is obviously not closed, but the adding the first moments once again makes it a closed system. This allows us to rewrite it into matrix form:

$$\frac{d}{dt}\begin{pmatrix} \langle g \rangle \\ \langle p \rangle \\ \langle g^2 \rangle \\ \langle gp \rangle \\ \langle p^2 \rangle \end{pmatrix} = \underbrace{\begin{pmatrix} \sigma_1 & v_2 & 0 & 0 & 0 \\ v_1 & \sigma_2 & 0 & 0 & 0 \\ \mu_g + v_1 & v_2 & 2\sigma_1 & 2v_2 & 0 \\ -v_1 & -v_2 & v_1 & \sigma_1 + \sigma_2 & v_2 \\ v_1 & \mu_p + v_2 & 0 & 2v_1 & 2\sigma_2 \end{pmatrix}}_{M} \begin{pmatrix} \langle g \rangle \\ \langle p \rangle \\ \langle g^2 \rangle \\ \langle gp \rangle \\ \langle p^2 \rangle \end{pmatrix}. \tag{4.10}$$

Since the eigenvalues of M are the eigenvalues of the upper left 2×2 block (M_1) and the eigenvalues of the lower right 3×3 block, and because the eigenvectors corresponding to the lower right block have zeros as first two entries, we can find study the second moments only from

$$M_2 = \begin{pmatrix} 2\sigma_1 & 2v_1 & 0 \\ v_1 & \sigma_1 + \sigma_2 & v_2 \\ 0 & 2v_1 & 2\sigma_2 \end{pmatrix}.$$

The eigenvalues of M_2 are $2\lambda_1$, $\lambda_1+\lambda_2$ and $2\lambda_2$ with eigenvectors respectively $\begin{pmatrix} (\frac{\sigma_1-\sigma_2+\Delta}{2v_1})^2 \\ \frac{\sigma_1-\sigma_2+\Delta}{2v_1} \\ 1 \end{pmatrix}$, $\begin{pmatrix} -\frac{v_2}{v_1} \\ \frac{\sigma_1-\sigma_2}{2v_1} \\ 1 \end{pmatrix}$ and $\begin{pmatrix} (\frac{\sigma_1-\sigma_2-\Delta}{2v_1})^2 \\ \frac{\sigma_1-\sigma_2-\Delta}{2v_1} \\ 1 \end{pmatrix}$. Thus the corresponding eigenvectors of M are these three vectors with two zeros added as first two entries. We are not as much interested in the remaining two eigenvectors of M, as these eigenvalues are lower so in the large time limit they will vanish. Like in the previous section, the eigenvalue $2\lambda_1$ has a strictly positive eigenvector by the Perron-Frobenius Theorem. Since the largest eigenvalue of M is $2\lambda_1$ and its corresponding eigenvector is $\begin{pmatrix} 0 \\ 0 \\ (\frac{\sigma_1-\sigma_2+\Delta}{2v_1})^2 \\ \frac{\sigma_1-\sigma_2+\Delta}{2v_1} \\ 1 \end{pmatrix}$, for $t \to \infty$ we get

$$\langle g^2 \rangle = \frac{\sigma_1 - \sigma_2 + \Delta}{2v_1} \langle gp \rangle = (\frac{\sigma_1 - \sigma_2 + \Delta}{2v_1})^2 \langle p^2 \rangle. \tag{4.11}$$

Unfortunately, it is not possible to simply find an expression like (4.7) for the second moments from M_2, as the second moments are dependent on the first moments through the remaining two eigenvectors of M. These eigenvectors are extremely complicated and thus it is not feasible to compute the second moments exactly. Although Lu et al. [25] claim to have done so, their result seems incorrect as it contains a factor $e^{3\lambda_1 t}$, which is impossible as the largest eigenvector is $2\lambda_1$, and their lack of mathematical methods makes error-checking impossible.

Note that the largest eigenvalue of M_2 is twice that of M_1. Because of this, in the noise $\eta_p = \frac{\langle p^2 \rangle - \langle p \rangle^2}{\langle p \rangle^2}$ the exponents will cancel for large t, and as such we expect it become constant. We show a proof of this in the next section. This result is very interesting, as it means that the variance and the square of the mean grow at the same exponential rate! We already saw that this holds for exponential growth in Section 2, and show a mathematical proof of this in the next section.

4.1.3. Time derivatives of noise and fraction

Alternative proof of 4.1. We showed before that the ratio $\frac{\langle g\rangle}{\langle p\rangle}$ converges to a steady state, which means that the fraction $\phi = \frac{\langle p\rangle}{\langle g\rangle+\langle p\rangle}$ does as well. Another way to find this result, is by using master equation and the expectation of the fraction $\langle\phi\rangle = \sum_{p,g} \frac{p}{p+g} P_{p,g}$ in a similar way to (4.4):

$$\begin{aligned}\frac{d}{dt}\langle\frac{p}{g+p}\rangle &= (\mu_p-\mu_g)\langle\frac{pg}{(p+g)(p+g+1)}\rangle + v_1\langle\frac{g}{g+p}\rangle - v_2\langle\frac{p}{g+p}\rangle \\ &\leq (\mu_p-\mu_g)\langle\frac{p}{g+p}\rangle\langle\frac{g}{g+p}\rangle + v_1\langle\frac{g}{g+p}\rangle - v_2\langle\frac{p}{g+p}\rangle \\ &= (\mu_p-\mu_g)\langle\phi\rangle(1-\langle\phi\rangle) + v_1(1-\langle\phi\rangle) - v_2\langle\phi\rangle \\ &= \langle\phi\rangle^2(\mu_g-\mu_p) + \langle\phi\rangle(\mu_p-\mu_g-v_1-v_2) + v_2.\end{aligned} \tag{4.12}$$

which gives the same value for ϕ as in (4.8) if we set this to zero, showing that $\langle\frac{p}{g+p}\rangle = \frac{\langle p\rangle}{\langle g\rangle+\langle p\rangle}$. Note that in the inequality we got $\leq$ by Jensen's inequality, as $\phi(1-\phi)$ is a concave function. This steady state of the fraction is stable, as shown by Bruggeman [27]. □

Given that the fraction converges, we also show that the noise converges by using the master equation. We also show that the noise in p and g is equal in this limit, as claimed in Theorem 4.2.

Proof of Theorem 4.2. Since $\eta_p = \frac{\langle p^2\rangle-\langle p\rangle^2}{\langle p\rangle^2}$, by the quotient rule and the product rule we get using (4.5) and (4.9)

$$\begin{aligned}\frac{d}{dt}\eta_p &= \frac{\langle p\rangle^2\frac{d\langle p^2\rangle}{dt} - \langle p^2\rangle\frac{d\langle p\rangle^2}{dt}}{\langle p\rangle^2} = \frac{\langle p\rangle^2\frac{d\langle p^2\rangle}{dt} - 2\langle p\rangle\langle p^2\rangle\frac{d\langle p\rangle}{dt}}{\langle p\rangle^2} \\ &= \frac{\mu_p\langle p\rangle^3 + v_1\langle p\rangle^2\langle g\rangle + v_2\langle p\rangle^3}{\langle p\rangle^4} + \frac{2v_1(\langle p\rangle^2\langle gp\rangle - \langle p^2\rangle\langle p\rangle\langle g\rangle)}{\langle p\rangle^4}.\end{aligned} \tag{4.13}$$

The first of these two terms clearly becomes zero for large t, as the numerator is order $\langle p\rangle$ smaller than the denominator, and $\langle p\rangle$ grows exponentially. The second term becomes zero if and only if

$$\frac{\langle p\rangle\langle gp\rangle - \langle p^2\rangle\langle g\rangle}{\langle p\rangle^3}. \tag{4.14}$$

goes to zero. If the fraction becomes constant, $p \approx \alpha g$ for some $\alpha \in \mathbb{R}$, so that this term becomes zero and the noise converges as well. In the that this indeed happens.

Another interesting observation is that for large t, $\langle g^2\rangle = (\frac{\sigma_1-\sigma_2+\Delta}{2v_1})^2\langle p^2\rangle$ as seen in (4.11). We already saw that $\langle g\rangle = \frac{\sigma_1-\sigma_2+\Delta}{2v_1}\langle p\rangle$ in this limit, but this means that

$$\eta_g = \frac{\langle g^2\rangle - \langle g\rangle^2}{\langle g\rangle^2} = \frac{(\frac{\sigma_1-\sigma_2+\Delta}{2v_1})^2\langle p^2\rangle - (\frac{\sigma_1-\sigma_2+\Delta}{2v_1})^2\langle p\rangle^2}{(\frac{\sigma_1-\sigma_2+\Delta}{2v_1})^2\langle p\rangle^2} = \frac{\langle p^2\rangle - \langle p\rangle^2}{\langle p\rangle^2} = \eta_p.$$

And thus the noise in g and p will be equal. □

As we showed before, the growing cells behave almost equal to exponential growth for low fractions. This means that if at the initial time $t = 0$ there are N_0 growing cells the noise in the growing cells will be approximately $\frac{1}{N_0}$, and as such the noise in the number of persisters will equal that. However, if the initial number of persister cells is high and the initial number of growing cells is low, this does not hold.

4.2. Dependence of noise on initial conditions

In Section 4.1 we found that the first and second moments of $P_{g,p}$ form a closed system of linear ordinary differential equations (4.10). Instead of solving this system analytically as before, we can numerically solve it using Matlab, which allows us to better visualize the results. Since we already showed that if $g > 0$ at $t = 0$ the number of persister cells have little influence on the noise, we are mostly interested in the scenario where we start with $g = 0$. Calculating the steady state noise with the same parameters as in Section 5.2 resulted in Figure 4.1[1]. Clearly, if $g > 0$ the noise immediately drops to about $\frac{1}{g}$, which is the noise in exponential growth. However, if $g = 0$ the noise is incredibly high for low persister numbers. It seems as if increasing the number of persister cells also has the effect of decreasing the steady state noise by a rate of $\frac{1}{p}$. This has a simple explanation: Given only persister cells, the noise is largely dependent on the first switch time. This time is an exponential distribution with mean $\frac{1}{v_2}$ for each cell, thus the first switch is an exponential distribution with mean $\frac{1}{pv_2}$. Thus increasing p decreases the first switch with rate $\frac{1}{p}$.

4.3. Exact model at constant persister fraction

If we assume the fraction of persister cells to be in steady state, i.e. either in the large time limit or by setting the initial conditions to already be such that the fraction is in steady state, we can compare the persister model to an exponential growth model. Note that the fraction will never be truly constant, but the expectation of the fraction will be. Since we know the expected value of ϕ from the previous section, we can now prove Theorem 4.3 using the results from Section 2.

Proof of Theorem 4.3. Suppose we have $N = g+p$ cells and a steady state fraction $\phi = \frac{p}{g+p}$. Then the growth rate μ_N of N will equal $\mu_g(1-\phi) + \mu_p\phi$. From (4.8) we know ϕ. We can rewrite it to get $\phi = \frac{\mu_g+v_1+v_2-\mu_p-\Delta}{2(\mu_g-\mu_p)}$. This yields for the growth rate:

$$\begin{aligned}\mu_N &= \mu_g + \phi(\mu_p - \mu_g) = \mu_g - \frac{1}{2}(\mu_g + v_1 + v_2 - \mu_p - \Delta) \\ &= \frac{1}{2}(\mu_g - v_1 + \mu_p - v_2 + \Delta) = \lambda_1.\end{aligned} \tag{4.15}$$

[1]Different parameters have different impact on the noise depending on the initial conditions. Most notably, v_2 has an enourmous impact on the noise, but only if the initial conditions contain no growing cells, which should make sense intuitively.

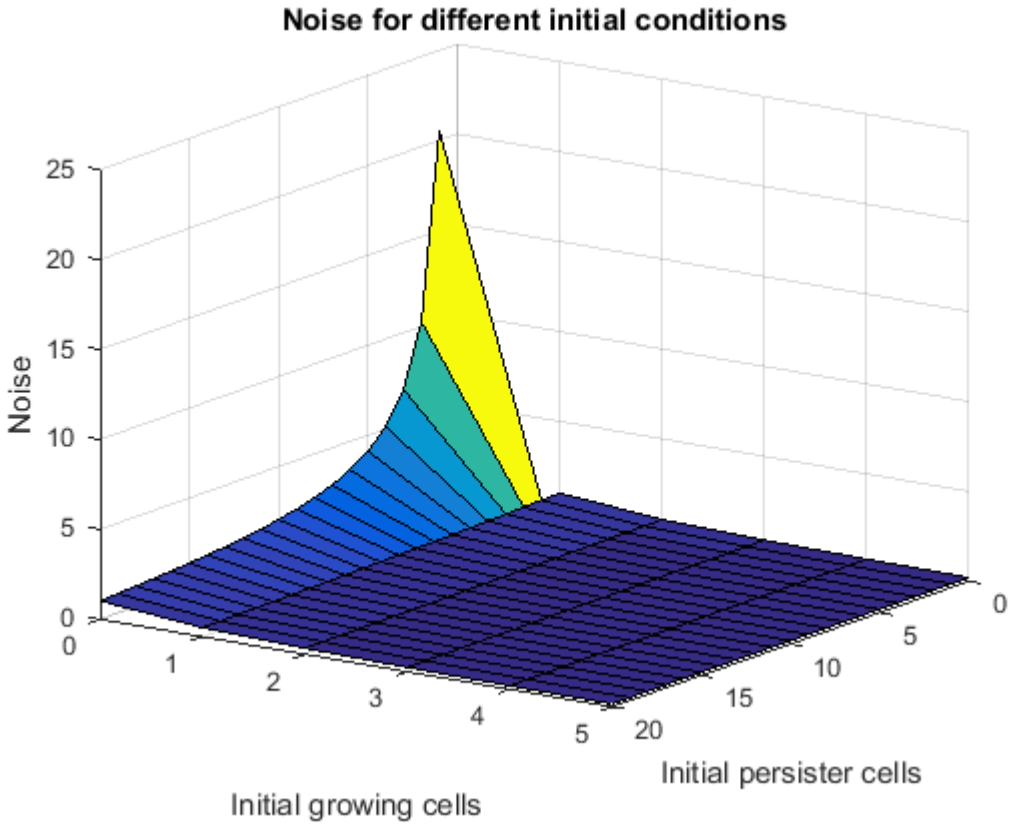


Figure 4.1.: Steady-state noise values for different initial conditions. Noise is about $\frac{1}{g}$ for initial condition $g > 0$. For $g = 0$ the noise increases massively, but seems to decrease with rate $\frac{1}{p}$. In this figure $\mu_g = 2$, $\mu_p = 0$, $v_1 = 0.12$ and $v_2 = 0.1$.

This is exactly as expected, as λ_1 is the growth rate as calculated in Section 4.1.1! However, λ_1 was shown to be the growth rate for large t, whereas here we only assumed the fraction to be in steady state, which could also be if we simply set the initial conditions to $g = \gamma p$. If we regard ϕ as a constant instead of a variable, this process is simply exponential growth with expectation $e^{\lambda_1 t}$. Thus the total amount of cells after time t follows a negative binomial distribution with parameters $N(0)$ and $q = e^{-\lambda_1 t}$. We are now interested in the number of persister cells given a certain number of total cells, as we know the probability distribution for the number of total cells. In Section 5.3 we show numerically that the distribution of persister cells given N total cells fits a binomial distribution with parameters (N, ϕ). Thus the total probability distribution is given by

$$\begin{aligned} P_{g,p}(t) &= P(N \text{ total cells}|t)P(p \text{ persister cells}|N \text{ total cells}) \\ &= \binom{N-1}{N(0)-1} q^{N(0)}(1-q)^{N-N(0)} \binom{N}{p} \phi^p (1-\phi)^{N-p}. \end{aligned} \tag{4.16}$$

Here $N(0)$ is the initial number of cells and $q = e^{-\lambda_1 t}$ with λ_1 as in (4.15). Note that the probability of N total cells and p persister cells is of course equal to the probability of g growing cells and p persister cells as $N = p + g$. □

Using Theorem 4.3 we can find the risk. Assuming $N(0) = 1$, setting $p = 0$ and summing over all N yields the following:

Corollary 4.4. *The risk of extinction in the constant fraction model is given by*

$$\begin{aligned} P(0 \text{ persisters}|t) &= \sum_{N=1}^{\infty} q(1-q)^{N-1}(1-\phi)^N \\ &= q(1-\phi)\sum_{N=0}^{\infty}[(1-q)(1-\phi)]^N \\ &= \frac{q(1-\phi)}{1-(1-q)(1-\phi)} \end{aligned} \tag{4.17}$$

$$= \frac{e^{-\frac{1}{2}(\mu_g - v_1 - v_2 + \mu_p + \sqrt{(\mu_g - v_1 - \mu_p + v_2)^2 + 4v_1 v_2})t}\left(1 - \left(\frac{2v_1}{\mu_g + v_1 - \mu_p + v_2 + \sqrt{(\mu_g - v_1 - \mu_p + v_2)^2 + 4v_1 v_2}}\right)\right)}{1 - \left(1 - e^{-\frac{1}{2}(\mu_g - v_1 - v_2 + \mu_p + \sqrt{(\mu_g - v_1 - \mu_p + v_2)^2 + 4v_1 v_2})t}\right)\left(1 - \left(\frac{2v_1}{\mu_g + v_1 - \mu_p + v_2 + \sqrt{(\mu_g - v_1 - \mu_p + v_2)^2 + 4v_1 v_2}}\right)\right)}.$$

This curve is shown in Figure 4.2 for the same parameter values as in Section 3.1.

As in Section 3.1, we are interested in how the reaction rates affect the risk. We again set $\mu_g = 2$ and $t = 1$. Figure 4.3 shows the same results as before: v_1 is the most impactful of the switching rates with respect to the risk. Figure 4.4 shows that again, for $1 \geq v_1 \geq 0.1$ the impact of v_2 is notable, but not as large as that of v_1.

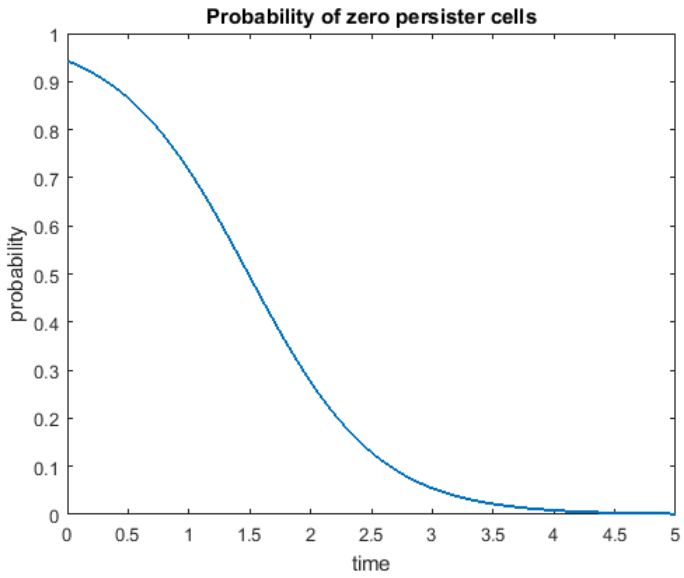


Figure 4.2.: Probability of zero persisters for the model with ϕ constant, as shown in (4.17). In this figure $\mu_g = 2$, $\mu_p = 0$, $v_1 = 0.12$ and $v_2 = 0.1$ so that $\phi \approx 0.057$.

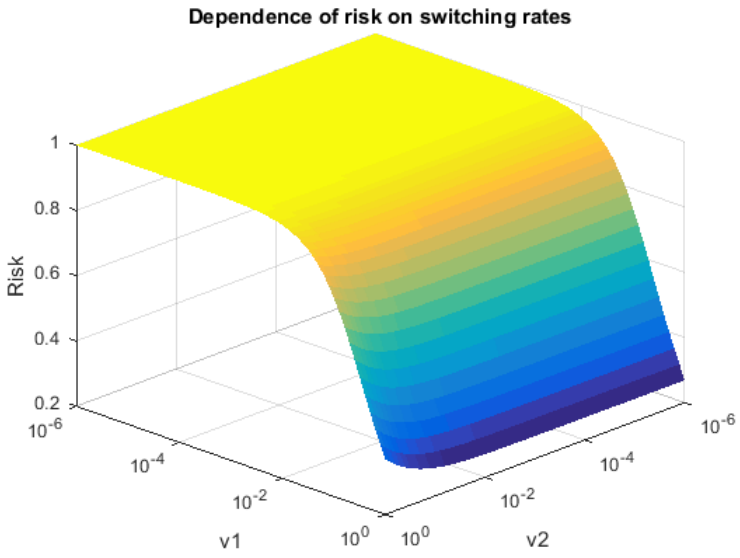


Figure 4.3.: Risk for $\mu_g = 2$ and $t = 1$ with different v_1 and v_2. Clearly, v_2 has very little influence on the risk whereas an increase in v_1 reduces the risk.

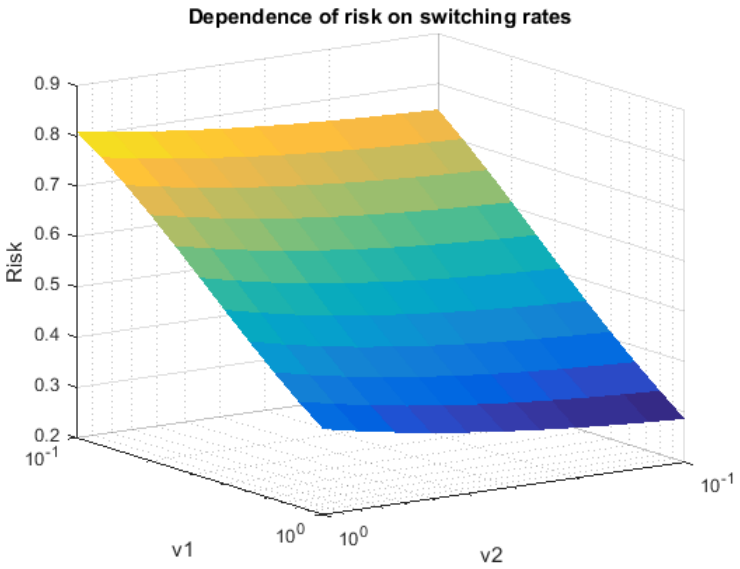


Figure 4.4.: Risk for $\mu_g = 2$ and $t = 1$ with different v_1 and v_2. This is a zoomed in version of Figure 3.2. This figure shows that for $v_1, v_2 \geq 0.1$ the impact of v_2 on the risk is notable.

5. Fitting distributions to data generated with the Gillespie Algorithm

Since we are unable to find an exact solution to the probability distribution of the persister model, we could try numerical methods to generate distributions. We can fit known probabilitty distributions to these generated histograms to validate our models. An exact method for numerically simulating the time evolution of a stochastic chemical process was introduced by Daniel Gillespie [18][19]. This method uses Monte Carlo techniques to simulate the Markov Process that is described by the master equation. Originally called the stochastic simulation algorithm (SSA), the Gillespie algorithm was designed for chemical reactions between molecules, but it can also be applied to other processes such as the model of persister cells, since the process is a Markov process. Extensions to non-Markovian processes have also been made, but these processes generally require significantly more computational power [30][31]. Using the Gillespie algorithm, we fit distributions of cell numbers over time in Section 5.2 and we fit the distribution of persister cells given a fixed total number of cells in Section 5.3.

5.1. Introduction to the Gillespie algorithm

A concise explanation of the framework of the Gillespie algorithm is given by Szallasi et al. [29]: Consider a system of N chemical species S_i, $i = 1, \ldots, N$ which interact in M different reactions R_j, $j = 1, \ldots, M$. The number of particles of each species at time t is denoted by $X_i(t)$. We want to study the evolution of the state vector $X(t) = (X_1(t), \ldots, X_N(t))^T$ over time, given initial conditions $X(0) = x_0$.

Each reaction R_j is characterized by two quantities. First, its state-change vector $\nu_j = (\nu_{1j}, \ldots, \nu_{Nj})$, which is the change to $X(t)$ when a reaction R_j occurs. Second, its propensity function a_j, which is defined such that, given $X(t) = x$, $a_j(x)dt$ is the probability that one R_j reaction occurs in the next infinitesimal time interval $[t, t + dt]$. As a_j is dependent on $X(t)$, it can change with time. For an unimolecular reaction such as $S_i \rightarrow$ *reaction products*, the propensity function can be seen as the rate of flow from S_i to its products. Naturally, this is dependent on the number of S_i particles, hence in this case $a_j(x) = c_j x_i$ where c_j is the reaction rate constant. For different reactions the propensity function will be different, but for our application to the persister model we have no need of those.

With this framework, the Gillespie algorithm can be defined in 5 steps:

1. Define the initial time t_0 and state $x = x_0$.
2. Evaluate all $a_j(x)$ for this t and x.
3. Generate values for the next reaction j and the time τ until this reaction (Monte Carlo step).
4. Replace x by $x + \nu_j$ and t by $t + \tau$.
5. If t is lower than some preset value *tstop*, repeat from step 2. Otherwise, end the simulation.

Gillespie suggested two different methods for the Monte Carlo step of the stochastic simulation algorithm: The 'first reaction method' and the direct method. In short, given a time t, population sizes x_i and reaction rates a_j, the direct method randomly generates the time τ until the next event and the event j. The first reaction method generates the tentative reaction time τ_j for each reaction, and then takes the minimum of all those τ_j and the corresponding reaction j. Clearly, the direct method requires 2 random numbers at each step, whereas the first reaction method requires M and thus is more computationally intensive. Thus, for $M > 2$ the direct method is preferred.

Since the process is assumed to be a Markov process, all reactions are exponentially distributed random variables with parameter a_j for reaction R_j. Thus for the first reaction method, we simply generate M samples from different exponential distributions and find the minimum. For the direct method, we only need to generate two random numbers. Because

$$\prod_j P(\tau_j > t) = \prod_j e^{-a_j t} = e^{-t\sum_j a_j}.$$

The minimum of multiple exponential distributions is an exponential distribution with parameter being the sum of all individual parameters. The time until the first event is thus an exponential random variable with parameter $\sum_{j=1}^{M} a_j$. The probability of this event being reaction R_j is $\frac{a_j}{a}$ where $a = \sum_j a_j$. By generating a random uniform number c on the interval $[0, a]$ we find the randomly generated reaction R_j by finding the smallest j for which $\sum_{k=1}^{j} a_k \geq c$. Thus by only taking two randomly generated numbers we find the time τ until the first reaction and the type of reaction R_j. Because of the Markov property, we can now repeat this step at time $t + \tau$.

5.2. Fitting distributions to cell data

We can now use the Gillespie algorithm to simulate the persister model defined by (1.1). Let $X(t) = (g(t), p(t))^T$ and R is the set of 4 reactions given in (1.1). Then $a_1 = \mu_g g(t)$, $a_2 =$

$\mu_p p(t)$, $a_3 = v_1 g(t)$ and $a_4 = v_2 p(t)$. We simulate many paths with the Gillespie algorithm with the goal of fitting the negative binomial distribution to the data. In Section 2 we have shown that exponential growth follows this distribution, and in Section 4.3 we show that it can be used as an approximation for the persister cell model. We fit distributions for different initial conditions for both the populations of growing and persister cells. The reaction rates used are $\mu_g = 2$, $\mu_p = 0$, $v_1 = 0.012$ and $v_2 = 0.1$. These rates were taken from Kussell et al. [2], with exception for v_1 which is reduced by a factor 10^4 to increase the fraction, as otherwise the simulations would be too computationally intensive. The fraction with these parameters is $\phi = 0.0057$ or $\gamma = 174$. For lower fractions, the model should resemble exponential growth even more, so this increase in fraction is justified. Unfortunately, results for fractions higher than this were inconclusive, as we were unable to generate enough data to do a goodness-of-fit test. Thus for populations of Hip cells where $\phi \approx 0.01$ these results do not hold.

There are two important things to note when interpreting the following figures: First, all fitted distributions are showed with a continuous red line. This is purely for clarity, as naturally the distributions used are discrete. Second, we fit distributions to the data to show that the data could have come from these distributions. We use goodness-of-fit χ^2 tests to show that we do not reject the hypothesis that the data is generated by the model. However, it is impossible to conclude that the data has been generated by the exact proposed model. A p-value higher than the significance level (usually 0.05) thus allows us only to conclude that the data could have been generated by the model.

If we take the initial conditions to be $g = 1$, $p = 0$, we expect a geometric distribution of the number of cells with parameter $e^{-\lambda_1 t} = e^{-\mu_g(1-\phi)t}$ as in section 4.3. We ran the Gillespie algorithm until $tstop = 3.5$, so that the expected number of cells is $e^{3.5\mu_g(1-\phi)} = e^{\lambda_1 t}$. In Figure 5.1 we see that the data (histogram) seems to fit the geometric distribution (red line). A goodness-of-fit χ^2 test yields a p-value of 0.2069, meaning the data fits the geometric distribution. If we change the initial conditions to $g = 20$, $p = 0$, we expect to see a sum of geometric distributions which is a negative binomial distribution. The expected parameters are of course $r = 20$ and $e^{-2\lambda_1}$, the number of geometric distributions and its original parameter. The fit is shown in Figure 5.2. The p-value of this fit is 0.5479, thus we conclude that the data fits the negative binomial distribution.

When setting the initial number of growing cells to zero, something interesting occurs. In Figure 5.3 we see that the distribution of the growing cell again fits a geometric distribution! The reason for this is that the persister cells do now grow themselves and v_2 is small compared to μ_g. Thus when the first persister cell switches to a growing cell, the process starts to grow with initial condition $g = 1$. The second switch of a persister cell to a growing cell usually occurs after the first cell has already grown to a larger population, thus the second switch has almost no influence on the growth. The average time until the first switch is e^{-20v_2}. However, if we try to fit a geometric distribution with parameter $e^{\lambda_1(t-e^{-20v_2})}$ we get a significantly low p-value. If we instead take as parameter the inverse of the mean of the data, the p-value is 0.0868, which is a much better fit. Thus the data fits a geometric distribution, but we do not know what the parameter is exactly.

However, we are of course interested in the behaviour of the parameter. We saw above

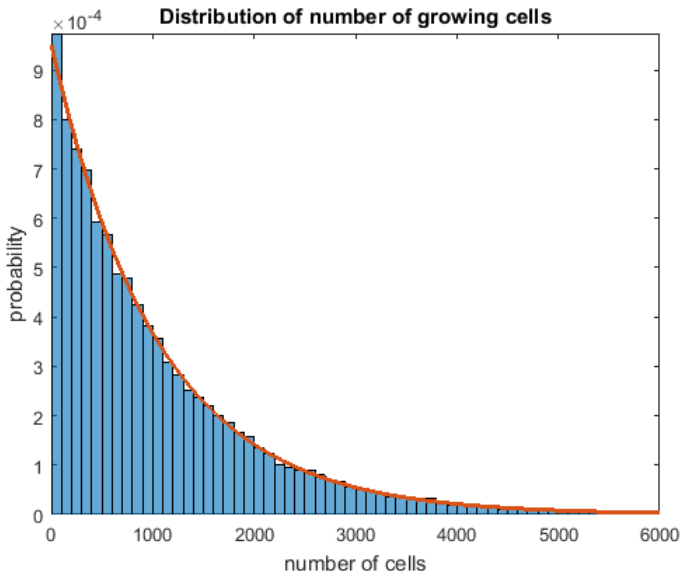


Figure 5.1.: Distribution of growing cells given initial condition $g = 1$, $p = 0$ and ratio $\gamma = 174.0$, fitted to a geometric distribution with parameter $e^{-\lambda_1 t}$.

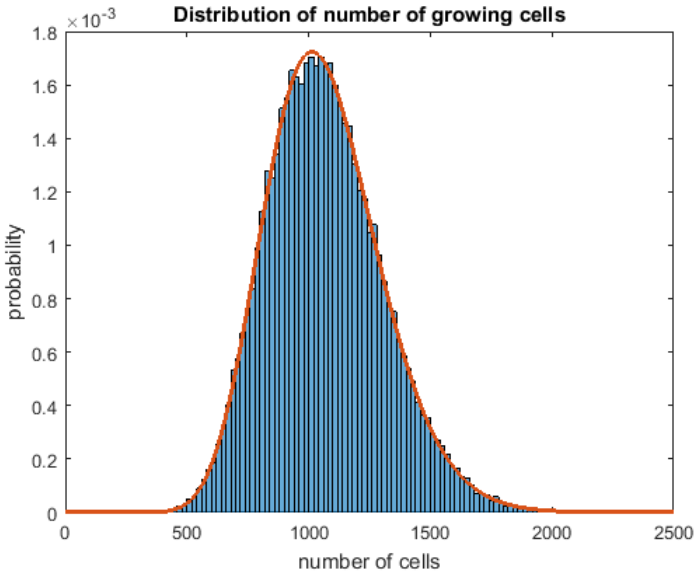


Figure 5.2.: Distribution of growing cells given initial condition $g = 20$, $p = 0$ and $\gamma = 174.0$, fitted to a negative binomial distribution with parameters 20 and $e^{-\lambda_1 t}$.

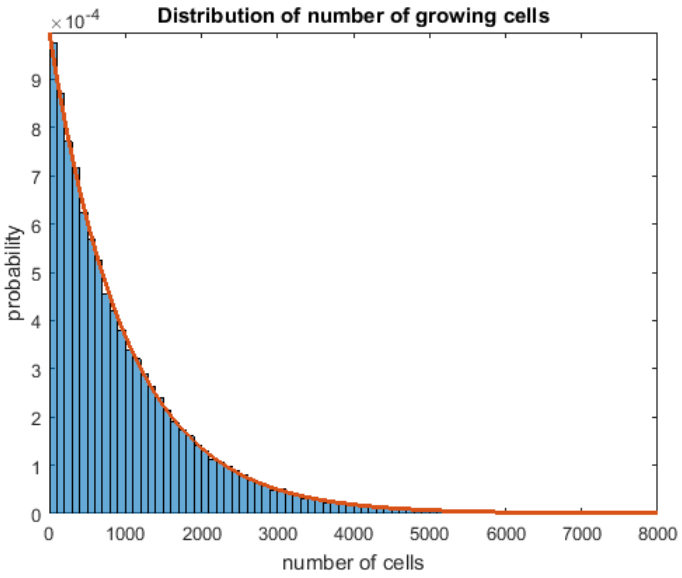


Figure 5.3.: Distribution of growing cells given initial condition $g = 0$, $p = 20$ and $\gamma = 174.0$. This looks a lot like the distribution of starting with a single growing cell, although they differ significantly. Fitted is a geometric distribution with mean determined empirically.

that simply treating the process as a lagged version of the process starting with a growing cell gives a lower mean than it should. This can be explained by two reasons: First, the first and second switch are not always far apart. If the second switch is close to the first, the expected number becomes much higher. Second, the time until the first switch is not always equal. Since the growth is exponential, outliers close to 0 have a massive influence, as they can grow to values multiple times as large as the expectation. Outliers on the other side do not cancel this out completely, as they still have positive expectation.

We also fitted distributions for persister cell numbers. As we expect the fraction to become constant, we expect persister cells to also fit a negative binomial distribution. For initial values $g = 1$, $p = 0$ and $tstop = 3.5$, we get Figure 5.4. We fit this to the geometric distribution with parameter $\frac{e^{-\lambda_1 t}}{\gamma}$. The resulting p-value is 0.2497, thus the data fits this geometric distribution. This also means that for initial conditions $g > 1$ the data should fit a negative binomial distribution.

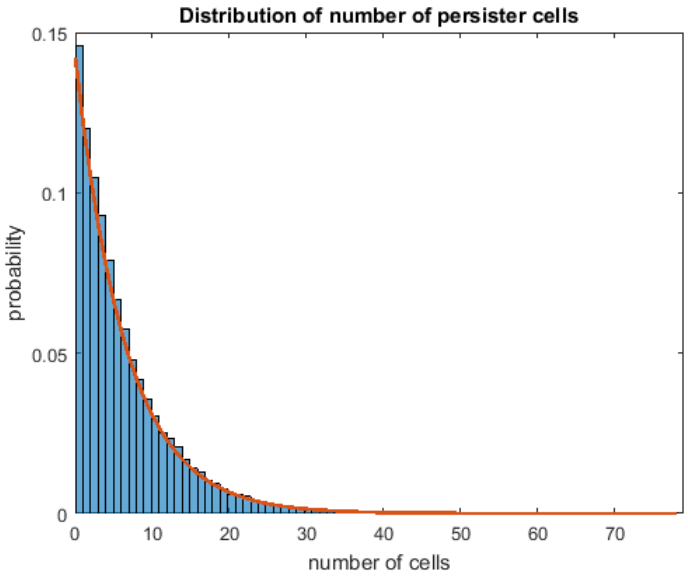


Figure 5.4.: Distribution of persister cells given initial condition $g = 1$, $p = 0$ and $\gamma = 174.0$, fitted to a geometric distribution with parameter$(\gamma^{-1}e^{-\lambda_1 t})$.

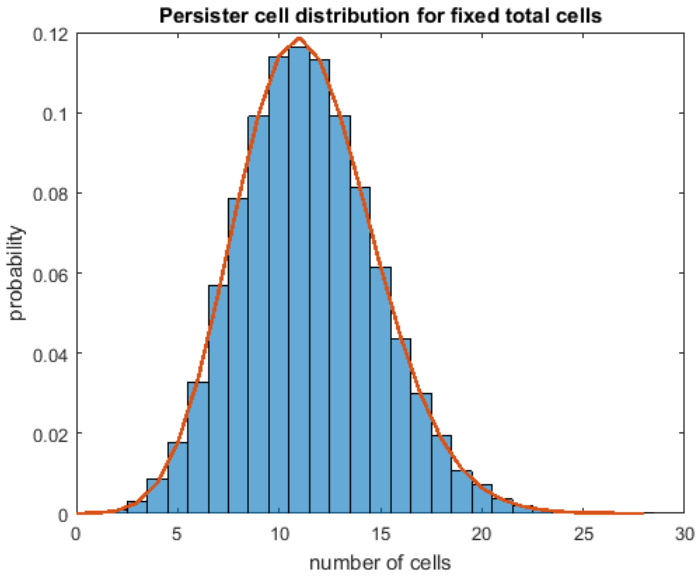


Figure 5.5.: Distribution of persister cells given $N = 2000$ and $\gamma = 174$, meaning the expected number of persisters is ≈ 11.5. The histogram is fitted to a binomial(2000,ϕ) distribution.

5.3. Distribution of persister cells for a given total population size

In Section 4.3 we showed that for a given constant fraction, the total number of cells follows a negative binomial distribution. If we would also know the distribution of persister cells given the total number of cells N and the fraction, we would be able to find an expression for $P_{g,p} = P(N \text{ total cells}|t)P(p \text{ persister cells}|N \text{ total cells})$. In this section we use a variation of the Gillespie algorithm to find such a distribution.

Instead of stopping the Gillespie algorithm by limiting the time as in step 5 of 5.1, we could run the algorithm but instead limit the total number of cells. We used the same parameters as in Section 5.2. As the most simple model, we expect a binomial distribution with parameters N and ϕ. Figure 5.5 shows that this is indeed a good fit, as the p-value is 0.2207. For low t the fraction usually is not yet in steady state, so we might expect to not fit a binomial distribution yet. However, Figure 5.6 shows that a binomial distribution is still a likely fit. Unfortunately, due to low computation power and extremely low probabilities for $p > 4$ as seen in the figure, we were unable to properly perform a goodness-of-fit test to this data.

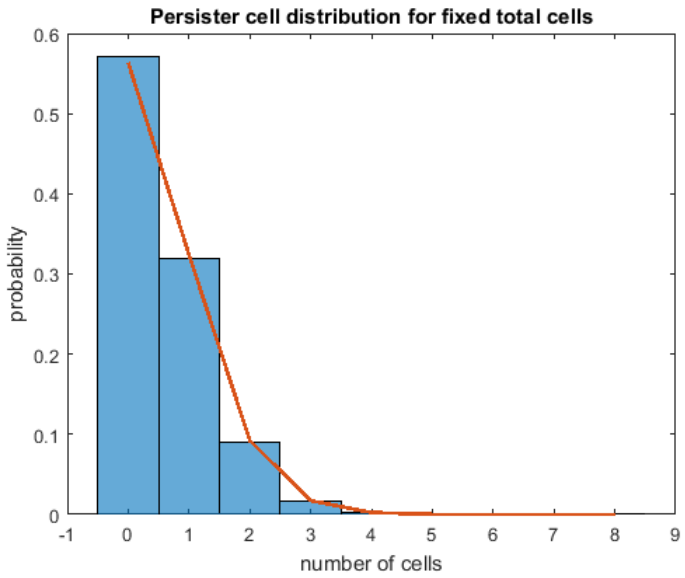


Figure 5.6.: Distribution of persister cells given $N = 100$ and $\gamma = 174$, meaning the expected number of persisters is ≈ 0.55. In red the binomial(100,ϕ) distribution is shown.

6. Results and Discussion

We aimed to study two questions related to the risk of extinction of a population: First, given initial conditions at $t = 0$, what is the risk at any $t > 0$? Second, given a population of certain size, what is the risk of this population? The answers to these questions depend on the probability distribution defined of the growing and persister cells. This distribution satiesfies the master equation (1.2). Although exact solutions to the master equation elude us still, we were able to find two accurate solutions to the risk of a population. Still, there is no simple answer to these questions, as we found that it depends on the initial conditions and the reaction rates.

First, we were able to find the exact solution to the generating function of a simplified model in 3.1. The exact expression for the risk in this model is given by (3.5). We showed this simplification to be accurate for persister fractions up to about 0.06 in Section 3.2, whereas usual fractions in cells are at most ≈ 0.01. For lower fractions, the approximation becomes more accurate thus this simplification can certainly be useful for applications.

Second, if we assumed the fraction of persister cells to be constant, we were also able to find an exact probability distribution for the model as shown in Section 4.3. This was possible by treating the process as an exponential growth process, of which we derived the exact probability distribution in Section 2. From this we derived another expression for the risk in this model, using the distribution of persister cells given the number of total cells, which we numerically showed to fit a binomial distribution in Section 5.3.

A comparison of the above two models is shown in Figure 6.1 and 6.2. One important note is that in both these figures we start with a single cell at time 0. If the number of cells at $t = 0$ would be n, the resulting graph would be the n'th power of these figures. This is because we assumed cells to grow independent of each other, thus the probability of zero persister cells in all processes is the product of having n times zero persister cells.

From Figure 6.1 we see that for high fractions, there are some differences. The main difference is in the initial conditions: In the simplified model we assume $g = 1, p = 0$ at time zero, whereas in the constant fraction model we start with $g = 1, p = 0$ with probability $1 - \phi$ and $g = 0, p = 1$ otherwise. This corresponds to knowing that we start with a single growing cell for the simplified model, and knowing we start with a single cell of unknown phenotype in the constant fraction model. However, for lower fractions the risk is almost equal in both models as we see in Figure 6.2. Thus for low fractions we can use (4.17) to approximate the risk of extinction even if at $t = 0$ know that we have a single growing cell! Compared to (3.5), this expression is much easier to work with and to exactly calculate, as it does not involve the integral present in the latter.

We also showed how the reaction rates influence the risk using (3.5) and (4.17). Naturally, both showed that an increase in μ_g decreases the risk, as a higher growth rate causes more

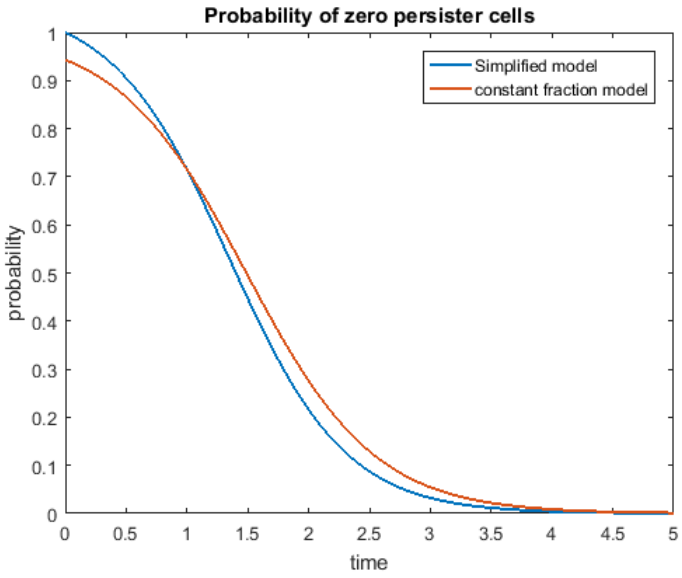


Figure 6.1.: Model comparison for the risk as derived in (3.5) and (4.17) with $\mu_g = 2$, $\mu_p = 0$, $v_1 = 0.12$ and $v_2 = 0.1$ so that $\gamma \approx 16.6$.

growing cells to be created, which causes more cells to be in the persister state. However, cells usually already try to maximize their growth rate, thus this is not nessecarily the most interesting observation. More interesting is how the switching rates affect the risk, as these rates are controlled by the cells. Both models also showed the same behaviour with respect to the switching rates v_1 and v_2, as seen in Figures 3.2 and 4.3. We see that v_1 has the largest impact on the risk of the two, which is to be expected as v_1 is responsible for creating persister cells. Since the number of persister cells is usually low, v_2 has little impact as the reverse switch rarely occurs. However, v_2 is more important if we were to consider populations that are exposed to stress, as it is responsible for restarting population growth when all growing cells have died. Unfortunately, we did not study this, but for further research in for example hedging strategies of persister cells, v_2 certainly plays an important role.

From the above, one might think that the best strategy for a cell is to simply maximize v_1. However, in the simplified model we assumed a switch from a growing cell to a persister cell to not consume a growing cell. As such, increasing v_1 would not be realistic. In the constant fraction model, increasing v_1 would increase ϕ and thus reduce the growth rate, which is not desired. Of course, the best strategy for a population of cells depends on the expected environment and the risk-growth rate tradeoff that they are willing to make.

Unfortunately, due to lacking computational power some results are not as elaborate as we would like. Mainly, we would like for the true model from Section 3.2 to show larger values of t. Also, in Section 5.2 we showed that the probability distributions of cell counts

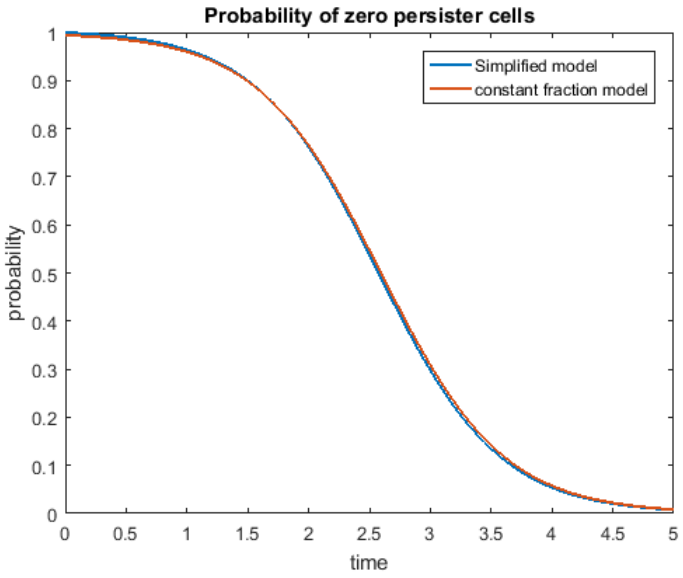


Figure 6.2.: Model comparison for the risk as derived in (3.5) and (4.17) with $\mu_g = 2$, $\mu_p = 0$, $v_1 = 0.012$ and $v_2 = 0.1$ so that $\gamma \approx 174$.

fit negative binomial distributions. However, we showed this for $\gamma \approx 174$, as for lower γ we were unable to generate enough data for a goodness-of-fit test to be accurate. For normal cell populations, $\gamma > 10^5$ so this will not cause any problems. However, for high-persistence cell populations $\gamma \approx 100$ in some cases. Thus for these populations we cannot yet fit their distributions. More research on this would increase our understanding of these distributions.

A fundamental assumption of our model was that all reaction waiting times are exponential random variables. However, in real-world applications these times are almost never exponential. Using non-exponential waiting times would cause the process to not be a Markov process, which is much harder to study mathematically. Kimmel and Axelrod [32] show some results for non-Markovian branching processes, but these processes are not nearly as well-studied as their Markovian counterparts. Simulating non-Markovian processes is a feasible alternative to avoid complex or possible impossible mathematics, but this is much more computationally intensive than simulating Markov processes. In the last few years some work on creating algorithms for non-Markovian processes has been done [30][31], which makes simulating non-Markovian processes feasible. However, no numerical solutions to the non-Markovian persister process have been studied and analytical solutions are also yet unexplored. We did construct a modified Gillespie algorithm to simulate the persister cell model for non-Markovian processes, but due to lack of computational power the results were inconclusive.

Some master equations are able to be solved exactly [28] by deriving a partial differential equation from the master equation. For our model it is possible to derive such an equation

as well, as we show in Section A.5. Unfortunately standard methods of solving this equation proved ineffective.

7. Popular summary

A well-known phenomenon in biology is bacterial resistance, in which bacteria adapt to become resistant to stress by mutating their DNA. Less known is bacterial persistence, in which cells exhibit a phenotypic switch to go into a dormant state, in which they are tolerant to stress. The main differences are that the phenotypic switch is reversible, and does not involve DNA mutations. Also, only a small subpopulation of bacteria is usually in this persister state. Whereas resistance usually decreases fitness in non-stressful environments, persisters have little effect on fitness. However, the rates at which cells switch their phenotype are random variables, thus the number of persister cells in a population varies heavily. Because of this, there is always a nonzero chance for a population to contain no persisters, thus being at risk of extinction. We aim to study this risk and related factors in this paper by constructing a mathematical model and studying the corresponding master equation, which is an equation that is equivalent to the model.

Since the fraction of persister cells is usually very low ($10^{-6} - 10^{-5}$), the model for persister cells is very much like that of a simple exponential growth process. For exponential growth, we derived an exact probability distribution. Using this, we constructed a simplified model for persister cells, which we were able to exactly solve. We have shown this simplified model to be accurate for low fractions of persister cells, thus having an accurate expression for the risk. From this expression we see that the growth rate is the most influencial parameter with regards to the risk, greatly reducing the risk if the growth rate is increased. However, cells usually already maximize their growth rate. Perhaps more interesting is the influence of the switching rates on the risk. We see that the switching rate from growing to persister cells has a major influence on the risk, whereas the reverse switching rate has very little influence. However, this parameter is still very influencial as it is responsible for the speed of population regrowth when all growing cells in a population die due to stress.

As persister cell numbers are hard to find experimentally, we used numerical methods to generate data on persister numbers. We used an algorithm called the Gillespie algorithm to exactly simulate the master equation. With this simulation, we generated many paths to accurately fit the probability distribution of the number of cells. Although this method provides no proof of the exact distribution, it does show the shape of the distribution to be close to what we expect on the basis of simple exponential growth.

Aside from the risk of the population, there are other factors that we are interested in. Most importantly, the expected fraction of persister cells and the noise (normalized variance) of the number of persisters. From the master equation, we could derive differential equations for both the first and second moments of the probability distribution, thus we were able to find the expectation and variance in the number of cells. From this we derived

an exact expression for the fraction of persister cells, and showed that both the fraction and the noise become constant in time. This is remarkable, as the number of cells keeps growing indefinitely. We also showed that the steady state noise depends very heavily on the initial conditions, and that the steady state fraction depends only on the reaction rates.

Bibliography

[1] Nathalie Q. Balaban, et al. (2004), *Bacterial persistence as a phenotypic switch* Science 305, 1622, DOI: 10.1126/science.1099390

[2] Edo Kussell, Roy Kishony, Nathalie Q. Balaban & Stanislas Leibler (1 April 2005), *Bacterial Persistence: A Model of Survival in Changing Environments* Genetics, vol. 169 no. 4 1807-1814; https://doi.org/10.1534/genetics.104.035352

[3] Joseph W. Bigger (14 Oct 1944), *Treatment of staphylococcal infections with penicillin*, The Lancet ii, pp 497

[4] Harris S. Moyed & Kevin P. Bertrand (1983), *hipA, a newly recognized gene of Escherichia coli K-12 that affects frequency of persistence after inhibition of murein synthesis*, Journal of bacteriology, vol 155, No2, p768-775

[5] K. Lewis, (2008), *multidrug tolerance of biofilm and persister cells*, Current Topics in Microbiology Immunology, 322, 107-31

[6] K. Lewis (October 2010), *Persister cells*, Annual Review of Microbiology, Vol. 64:357-372

[7] Clinton C. Dawson, Chaidan Intapa & Mary Ann Jabra-Rizk (2011), *Persisters: Survival at the Cellular Level*, PLoS Pathog 7(7): e1002121. https://doi.org/10.1371/journal.ppat.1002121

[8] Kyle R. Allison, Mark P. Brynildsen & James J. Collins (Oct 2011), *Heterogeneous Bacterial Persisters and Engineering Approaches to Eliminate them*, Current Opinions in Microbiology, 14(5): 593598

[9] K. Lewis (2000), *Programmed death in bacteria*, Microbiology and Molecular Biology Reviews, 64(3):503-14.

[10] R. Jayaraman (Dec 2008), *Bacterial persistence: some new insights into an old phenomenon*, Journal of Biosciences, 33(5):795-805

[11] Thomas K. Wood, Stephen J. Knabel & Brian W. Kwana (2013), *Bacterial Persister Cell Formation and Dormancy*, Applied and Environmental Microbiology, vol. 79 no. 23, 7116-7121

[12] Robert A. Fisher, Bridget Gollan & Sophie Helaine (May 2017), *Persistent bacterial infections and persister cells*, Nature Reviews Microbiology, doi:10.1038/nrmicro.2017.42

[13] R.C. Massey, A. Buckling & S.J. Peacock (13 Nov 2001), *Phenotypic switching of antibiotic resistance circumvents permanent costs in Staphylococcus aureus*, Current Biology 11(22):1810-4.

[14] Carl T. Bergstrom & Michael Lachmann (2004), *Shannon Information and Biological Fitness*, Information Theory Workshop, IEEE

[15] Bernadett Gaal, Jonathan W. Pitchford & A. Jamie Wood (April 2010) *Exact Results for the Evolution of Stochastic Switching in Variable Asymmetric Environments*, Genetics 109.113431.

[16] N. G. Van Kampen (21 March 2007), *Stochastic Processes in Physics and Chemistry*, North Holland, ISBN 9780444529657

[17] Donald A. McQuarrie (1967), *Stochastic approach to chemical kinetics* Journal of Applied Probability 4, 413-478

[18] Daniel T. Gillespie (December 1976), *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, Journal of Computational Physics Volume 22, Issue 4, Pages 403-434

[19] Daniel T. Gillespie (1977), *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem., 81 (25), pp 23402361, DOI: 10.1021/j100540a008

[20] W.H. Furry (15 September 1937), *On fluctuation phenomena in the passage of high energy electrons through lead*, Physical Review vol. 52, 569-581

[21] Joe Blitzstein on quora.com (2014), https://www.quora.com/Why-is-the-instantaneous-distribution-of-a-discrete-exponential-growth-process-described-by-a-negative-binomial-distribution

[22] David G. Kendall (1949), *Stochastic processes and population growth*, Journal of the royal statistical society, series B, vol. 11, no. 2, 230-282

[23] Crispin W. Gardiner (1985), *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* second edition, Springer, ISBN 3-540-61634-9

[24] Andrei D. Polyanin & Valentin F. Zaitsev (2002), *Handbook of exact solutions for ordinary differential equations*, Chapman&Hall/CRC

[25] Ting Lu, Tongye Shen, Matthew R. Bennett, Peter G. Wolynes & Jeff Hasty (27 November 2007), *Phenotypic variability of growing cellular populations*, PNAS vol. 104, no 48., 18982 - 18987

[26] Roger A. Horn & Charles R. Johnson (2013), *Matrix Analysis, second edition*, Cambridge University Press

[27] Frank J. Bruggeman, *Single cell biophysics* (unpublished)

[28] V. Shahrezaei, P. S. Swain (November 11, 2008), *analytical distributions for stochastic gene expression*, PNAS vol. 105 no. 45, 1725617261

[29] Z. Szallassi, J Stelling & V, Periwal (2006), *System Modelling in Cellular Biology: From concepts to nuts and bolts*, ISBN 0-262-19548-8

[30] N. Masuda & L. E. C. Rocha (24 Aug 2016), *A Gillespie algorithm for non-Markovian stochastic processes*, arXiv:1601.01490v2 [physics.soc-ph]

[31] M. Boguna, L.F. Lafuerza, R. Toral & M. A. Serrano (30 Jul 2014), *Simulating non-Markovian stochastic processes*, arXiv:1310.0926v3 [cond-mat.dis-nn]

[32] Marek Kimmel & David E. Axelrod (2000) *Branching Processes in Biology* Springer, ISBN 0-387-95340-X

[33] Theodore E. Harris (1963) *The Theory of Branching Processes* Springer-Verlag

[34] MATLAB 2016a, The MathWorks, Natick, 2016

A. Appendix

A.1. Markov property of models with exponential waiting times

Suppose we have a process Y with (not necessarily finite nor countable) states y and let $P(y_i, t_i)$ be the probability of the process being in state y_i at time t_i. This process is a Markov process if the equality

$$P(y_n, t_n | y_1, t_1; \ldots; y_{n-1}, t_{n-1}) = P(y_n, t_n | y_{n-1}, t_{n-1}). \tag{A.1}$$

holds for all for any set $t_1 < t_2 < \ldots < t_n$. This means that the future of the process is independent of all past states, thus only dependent on the currect state.

To be able to validate the master equation in the next section for our model, we have to prove that it follows a Markov process. We show that any multitype process of which the rates are distributed exponentially is a Markov Process. The state space of multitype process with n types is $\mathbb{Z}^n_{\geq 0}$. We will show that the transition probabilities satisfy (A.1). We know that for the exponential distribution with parameter λ the probability of the hitting time being larger than some t is

$$P(T > t) = e^{-\lambda t}.$$

The probability of multiple exponential distributions having hitting time larger than t is given by the probability of all their individual hitting times being larger than t, thus the product of the probabilities

$$\prod_i P_i(T_i > t) = \prod_i e^{-\lambda_i t} = e^{-t \sum_i \lambda_i}.$$

which is also an exponential distribution with parameter $\sum_i \lambda_i$. We will show that this means that the model satisfies (A.1). For an exponential distribution with parameter λ, it holds for $t > 0$ by the definition of conditional probability

$$P(T > s + t | T > s) = \frac{P(T > t + s)}{P(T > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t).$$

This means that the probability of an event occurring in the interval $[0, t]$ is equal to the probability of an event occuring in the interval $[s, s+t]$. This implies that the future states depend only on the currect state, thus (A.1) is satisfied.

A.2. Derivation of the master equation

For a source and more elaboration on this section see Van Kampen(2007) [16].

Suppose a process $P(y_i, t_i)$ satisfies (A.1). This allows us to compute probabilities of events using only the initial probability at time t_1 and transition probabilities:

$$P(y_1, t_1; y_2, t_2; y_3, t_3) = P(y_1, t_1)P(y_2, t_2|y_1, t_1)P(y_3, t_3|y_2, t_2). \tag{A.2}$$

Integrating this with respect to y_2 yields

$$P(y_1, t_1; y_3, t_3) = P(y_1, t_1) \int P(y_2, t_2|y_1, t_1)P(y_3, t_3|y_2, t_2)dy_2. \tag{A.3}$$

We then use $P(y_1, t_1; y_3, t_3) = P(y_1, t_1)P(y_3, t_3|y_1, t_1)$ and divide by $P(y_1, t_1)$

$$P(y_3, t_3|y_1, t_1) = \int P(y_2, t_2|y_1, t_1)P(y_3, t_3|y_2, t_2)dy_2. \tag{A.4}$$

This is a well-known result, better known as the Chapman-Kolmogorov equation. If we rewrite $T_\tau(y_2|y_1) = P(y_2, t_2|y_1, t_1)$, where $\tau = t_1 - t_2$, equation A.4 becomes

$$T_{\tau+\tau'}(y_3|y_1) = \int T_{\tau'}(y_3|y_2)T_\tau(y_2|y_1)dy_2, \tag{A.5}$$

where $\tau' = t_3 - t_2$.

Now let τ' be small. We look more closely at the probability $T_{\tau'}(y_3|y_2)$. Let $W(y_3|y_2)$ be the transition probability per unit time from y_2 to y_3. Clearly, the total rate of leaving state y_3 is given by $a_0(y_3) = \int W(y_2|y_3)dy_2$. Hence, the probability of staying in state y_3 for time τ' is $1 - a_0(y_3)\tau'$. Note that this assumes a maximum of one transition during time τ'. Furthermore, the probability of directly transitioning (without intermediate states) from state y_2 to y_3 in time τ' is given by $\tau' W(y_3|y_2)$. Again, this assumes τ' to be small enough to allow only one transition. Lastly, the probability of moving from state y_2 to a different state twice is $O(\tau'^2)$. Thus, the total probability of transitioning from state y_2 to state y_3 in time τ' is given by

$$(1 - a_0(y_3)\tau'\delta(y_3 - y_2) + \tau' W(y_3|y_2) + O(\tau'^2). \tag{A.6}$$

Here δ denotes the δ-distribution. Since we assumed τ' to be small, we can disregard the last term. If we insert this in the Chapman-Kolmogorov equation this becomes

$$\begin{aligned} T_{\tau+\tau'}(y_3|y_1) &= \int ((1 - a_0(y_3)\tau')\delta(y_3 - y_2) + \tau' W(y_3|y_2)T_\tau(y_2|y_1)dy_2 \\ &= (1 - a_0(y_3)\tau')T_\tau(y_3|y_1) + \tau' \int W(y_|y_2)T_\tau(y_2|y_1). \end{aligned}$$

Dividing by τ' and rewriting gives

$$\frac{T_{\tau+\tau'}(y_3|y_1) - T_\tau(y_3|y_1)}{\tau'} = -a_0(y_3)T_\tau(y_3|y_1) + \int W(y_|y_2)T_\tau(y_2|y_1). \tag{A.7}$$

Taking the limit $\tau' \to 0$ and writing out $a_0(y_3)$ yields

$$\frac{dT_\tau(y_3|y_1)}{dt} = \int W(y_3|y_2)T_\tau(y_2|y_1) - W(y_2|y_3)T_\tau(y_3|y_1)dy_2. \tag{A.8}$$

This is known as the master equation. By removing the conditioning on y_1, it is more commonly known in simpler form:

$$\frac{\partial P(y,t)}{\partial t} = \int W(y|y')P(y',t) - W(y'|y)P(y,t)dy'. \tag{A.9}$$

Or for discrete state spaces, which we will use later:

$$\frac{\partial P_n(t)}{\partial t} = \sum_n W_{nn'}P_{n'}(t) - W_{n'n}P_n(t). \tag{A.10}$$

Where $W_{nn'}$ is the rate of transition from state n' to state p.

A.3. Non-Exponential waiting times generate Non-Markovian processes

It is of importance to note that non-exponential waiting times would lead to a process that is not Markovian. Because of this, studying such processes is much harder, as the master equation does not hold. We will show that the only continuous probability distribution that has this property is in fact the exponential distribution. If we assume a process to be memoryless, i.e.

$$P(T > t + s|T > t) = P(T > s),$$

and by using the definition of conditional probability we get

$$P(T > t + s) = P(T > s)P(T > t).$$

This gives the equation

$$G(t + s) = G(t)G(s).$$

The function G now satisfies

$$\begin{aligned} G(2) &= G(1)^2 \\ G(3) &= G(2)G(1) = G(1)^3 \\ &\vdots \\ G(n) &= G(1)^n. \end{aligned}$$

Also,

$$\begin{aligned} G(1) &= G(\frac{1}{2})^2 \\ G(\frac{1}{2}) &= G(\frac{1}{4})^2 \\ &\vdots \\ G(\frac{1}{2^n}) &= G(\frac{1}{2^{n+1}})^2. \end{aligned}$$

so that $G(\frac{1}{2^n}) = G(1)^{\frac{1}{n}}$ Then taking the binomial expansion $\frac{p}{q} = \sum_{n=1}^{\infty} a_n 2^{-n}$ we get for $\frac{p}{q} \in (0,1)$

$$G(\frac{p}{q}) = G(\sum_{n=1}^{\infty} a_n 2^{-n}) = \prod_{n=1}^{\infty} G(a_n 2^{-n}) = \prod_{n=1}^{\infty} G(1)^{a_n 2^{-n}} = G(1)^{\prod_{n=1}^{\infty} a_n 2^{-n}} = G(1)^{\frac{p}{q}}.$$

Here $a_n \in \{0,1\}$. We will not prove that all rational numbers in $(0,1)$ have a unique binomial expansion, as it is a well-known result.

This relation means that

$$G(x) = G(1)^x = e^{\log(G(1))x} = e^{-\lambda x}.$$

for all $x \in \mathbb{Q}_{\geq 0}$. Since $\mathbb{Q}$ lies dense in $\mathbb{R}$, the relations holds for all $x \in \mathbb{R}_{\geq 0}$. Since $\lambda = G(1) \geq 0$ as G is a probability, any memoryless distribution must be an exponential distribution.

A.4. Probability distributions

A.4.1. Binomial distribution

The binomial distribution is a discrete probability distribution with parameters $p \in [0,1]$ and $n \in \mathbb{N}$. It has probability mass function

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Its mean and variance are np and $np(1-p)$.

A.4.2. Geometric distribution

The geometric distribution is a discrete probability distribution with parameter $p \in [0,1]$ and probability mass function

$$P(X = k) = (1-p)^{k-1} p.$$

And generating function

$$F(z) = \frac{pz}{1-(1-p)z}.$$

Its mean and variance are $\frac{1}{p}$ and $\frac{1-p}{p^2}$ respectively.

A.4.3. Negative binomial distribution

The negative binomial distribution is a sum of r independent geometric distributions with parameter $p \in [0,1]$. It has probability mass function

$$P(X=k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}.$$

And generating function

$$F(z) = (\frac{pz}{1-(1-p)z})^r.$$

Its mean and variance are $\frac{r}{p}$ and $\frac{r(1-p)}{p^2}$ respectively.

A.4.4. Exponential distribution

The exponential distribution is a continuous probability distribution with the special property of memorylessness. It has parameter $\lambda > 0$ and probability density function

$$P_X(x) = \lambda e^{-\lambda x}.$$

Its mean and variance are $\frac{1}{\lambda}$ and $\frac{1}{\lambda^2}$ respectively. It is a special case of the Gamma distribution.

A.4.5. Gamma distribution

The Gamma distribution is a continuous probability distribution with shape and rate parameters $\alpha > 0$ and $\beta > 0$ and probability density function

$$P_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}.$$

Its mean and variance are $\frac{\alpha}{\beta}$ and $\frac{\alpha}{\beta^2}$ respectively. For $\alpha = 1$ this is an exponential distribution.

A.5. PDE methods for solving the master equation

As discussed in previous sections, the generating function (1.3) and the probability mass function defined by the master equation (1.2) are equivalent. Hence, a solution to one admits a solution to the other. By using the master equation, we can find a partial

differential equation (pde) for the generating function. Although these pde's are generally not solvable, in some specific cases there are solutions [28].

Combining (1.2) and (1.3) yields

$$\begin{aligned}\frac{\partial F}{\partial t} &= \sum_{g,p} s_1^g s_2^p \frac{dP_{g,p}(t)}{dt} \\ &= \sum_{g,p} s_1^g s_2^p \{\mu_g[(g-1)P_{g-1,p} - gP_{g,p}] + \mu_p[(p-1)P_{g,p-1} - pP_{g,p}] \\ &+ v_1[(g+1)P_{g+1,p-1} - gP_{g,p}] + v_2[(p+1)P_{g-1,p+1} - pP_{g,p}]\}.\end{aligned} \tag{A.11}$$

Since $P_{g,p} = 0$ for $g < 0$ or $p < 0$, for the μ_g term the following holds:

$$\begin{aligned}\mu_g \sum_{g,p} s_1^g s_2^p[(g-1)P_{g-1,p} - gP_{g,p}] &= \mu_g \sum_{g,p} s_2^p s_1^g g(s_1-1)P_{g,p} \\ &= \mu_g s_1(s_1-1)\sum_{g,p} g s_1^{g-1} s_2^p P_{g,p} \\ &= \mu_g s_1(s_1-1)\frac{\partial F}{\partial s_1}.\end{aligned}$$

Similar equations can be done for the other three terms in (A.11). This together yields a partial differential equation for the generating function:

$$\begin{aligned}\frac{\partial F}{\partial t} &= \mu_g s_1(s_1-1)\frac{\partial F}{\partial_g} + \mu_p s_2(s_2-1)\frac{\partial F}{\partial s_2} + v_1(s_2-s_1)\frac{\partial F}{\partial s_1} + v_2(s_1-s_2)\frac{\partial F}{\partial s_2} \\ &= \frac{\partial F}{\partial s_1}[\mu_g s_1(s_1-1) + v_1(s_2-s_1)] + \frac{\partial F}{\partial s_2}[\mu_p s_2(s_2-1) + v_2(s_1-s_2)].\end{aligned} \tag{A.12}$$

Unfortunately, despite the symmetry in this equation it is not solvable. We could numerically solve (A.12) using e.g. finite difference methods. However, we are interested not only in the pgf but also in all of its derivatives. Since numerical methods become less accurate for higher-order derivates, we refrain from trying to find solutions this way and instead decided to focus on other approaches. However, if we study exponential growth, i.e. we set $v_1 = v_2 = \mu_p = 0$, we get

$$\frac{\partial F}{\partial t} = \frac{\partial F}{\partial s_1}\mu_g s_1(s_1-1).$$

This equation was first solved by David Kendall in 1949 [22] and indeed has solution (2.3).