

A stochastic model of cell cycle desynchronization

Peter Olofsson* and Thomas O. McDonald†

November 5, 2009

Abstract

A general branching process model is suggested to describe cell cycle desynchronization. Cell cycle phase times are modeled as random variables and a formula for the expected fraction of cells in S phase as a function of time is established. The model is compared to data from the literature and is also compared to previously suggested deterministic and stochastic models.

Keywords: Cell cycle, branching process, desynchronization, stable phase distribution *AMS 2000 Mathematics Subject Classification Codes:* 60G99, 60K99, 62P10, 92D25.

1 Introduction

The cell cycle of a eukaryotic cell consists of four phases. In brief, they are G_1 where the cell grows, S where its DNA synthesizes, G_2 where it prepares for division, and M where it divides. Due to variability in individual cell phase times, an initially synchronous population will eventually lose synchronicity and the percentages of cells in the difference phases settle in toward a stable phase distribution. Questions of interest to biologists and mathematical

*Corresponding author. Trinity University, Mathematics Department, One Trinity Place, San Antonio, TX 78212. Email: polofsso@trinity.edu

†Rice University, Department of Statistics, 6100 Main Street, Houston, TX 77005

modelers alike include how the stable distribution relates to phase time parameters and how, and at what rate, the phase percentages approach the stable distribution.

Some early work on stable distributions in cell kinetics was done by Bronk *et al.* 1968, Bronk 1979, and Macdonald, 1970, 1974, 1978, and 1981. In fact, some of the stable population results that we prove below existed already in Macdonald's papers with mostly heuristic derivations. An important publication on the issue is Chiorino *et al.* (2001). There, a deterministic model is developed and fitted to data from various cell lines, obtained specifically for this purpose. A system of partial differential equations is established and investigated via an asymptotic approximation of the solution. One noticeable feature is the oscillatory pattern in which cell phase percentages approach the stable phase distribution, a pattern seen in the data as well as in the solution to the model. The authors focus mainly on cells in S phase, starting by labeling cells in that phase, in a cell population in stable exponential growth, and then measuring the fraction of cells in S phase at regular time intervals. By design, initially 100% of cells are in S phase and as time advances, the percentages oscillate to settle in toward a stable limit. The model is fitted to the data and approximate relations between phase time parameters on the one hand and the stable phase distribution together with convergence rate and periodicity on the other are established. The data from Chiorino *et al.* was also used in Milotti *et al.* (2008) where a nice heuristic stochastic model was suggested, based in part on Bronk *et al.* (1968).

We propose a general branching process model. Thus, our approach is stochastic, putting us closer to the second of the papers mentioned above. Our model more faithfully describes the biological processes and incorporates sampling effects due to exponential growth that are disregarded by Milotti *et al.* In the model, we describe phase times as random variables, consider a population that reproduces by splitting, and keep track of the number of cells that are in S phase at any given time, as well as the total number of cells. For ease of reading, in the next section we give a brief description of the general branching process (or Crump-Mode-Jagers process). For a comprehensive treatment, see Jagers and Nerman (1984).

2 General branching processes

The fundamental mathematical object in a general branching process is the *reproduction process*, ξ . This is a stochastic point process on $[0, \infty)$ that describes how an individual reproduces, thus $\xi(a) = \int_0^a \xi(dt)$ gives the number of children up to age a . Each newborn individual starts reproducing according to a copy of ξ , independently of other individuals. In cell populations where reproduction is by division, ξ is characterized by two random variables: the lifetime L and the number of offspring X . Specifically

$$\xi(dt) = X\delta_L(dt)$$

where δ_L is the unit point mass at L (L and X may be dependent). The expression for ξ simply means that the cell lives for a time L , then produces X daughter cells. If there is no death, $X \equiv 2$.

To capture the growth rate of the process, we consider the *mean reproduction process*, $\mu(dt) = E[\xi(dt)]$, which thus defines a measure. Of particular interest is its Laplace transform

$$\hat{\mu}(r) = \int_0^\infty e^{-rt} \mu(dt)$$

The growth rate is now determined by the *Malthusian parameter* which is the number α that satisfies $\hat{\mu}(\alpha) = 1$. If $\alpha > 0$, the process is said to be *supercritical* which means that the population has a chance of avoiding extinction and that the expected population size grows asymptotically as $e^{\alpha t}$. We deal with cell populations with no death so our process is supercritical and the equation defining α becomes

$$2\hat{F}(\alpha) = 1 \tag{2.1}$$

where F is the distribution function of L and $\hat{F}(\alpha) = \int_0^\infty e^{-\alpha t} F(dt)$. If there is death, the number 2 is replaced by the mean number of surviving daughter cells.

To count, or measure, the population, *random characteristics* are used. A random characteristic is a real valued stochastic or deterministic process χ , where $\chi(a)$ gives the contribution of an individual of age a . We assume that χ is nonnegative and vanishing for negative a (no individual contributes before her birth). In a general branching process, individuals are identified by descent; the ancestor is denoted by 0, the individual $x = (x_1, \dots, x_n)$ belongs

to the n th generation and is the x_n th child of the x_{n-1} th child of...of the x_1 th child of the ancestor. This gives the *population space*

$$I = \bigcup_{n=0}^{\infty} N^n,$$

that is, the set of all conceivable individuals (whether an individual is actually born depends on the reproduction process of its mother). Note that this construction assumes that we start the population from one individual, the ancestor, an assumption we make throughout this paper.

Now let the χ -value pertaining to the individual x be denoted by χ_x and denote the birth time of the individual x by τ_x . The χ -counted population, Z_t^X is defined as

$$Z_t^X = \sum_{x \in I} \chi_x(t - \tau_x),$$

the sum of the contributions of all individuals (at time t the individual x is of age $t - \tau_x$). The simplest example of a random characteristic is the indicator function $\chi(t) = I\{t \geq 0\}$, which equals zero before the individual is born and one thereafter, in which case Z_t^X is the number of individuals born up to time t .

We next state without conditions the main convergence results that we need in this paper. The results can be found in Jagers and Nerman (1984) and for convenience, we also state them with all their conditions in Appendix A. Let

$$E[\hat{\chi}(\alpha)] = \int_0^{\infty} e^{-\alpha t} E[\chi(t)] dt$$

and

$$\beta = 2 \int_0^{\infty} t e^{-\alpha t} F_L(dt) \tag{2.2}$$

Then

$$E[e^{-\alpha t} Z_t^X] \rightarrow \frac{E[\hat{\chi}(\alpha)]}{\beta} \tag{2.3}$$

as $t \rightarrow \infty$, and also

$$e^{-\alpha t} Z_t^X \rightarrow \frac{E[\hat{\chi}(\alpha)]}{\beta} W \tag{2.4}$$

almost surely as $t \rightarrow \infty$, where W is a random variable with mean 1. In the present paper, our main use of the theorem is to establish asymptotic proportions of cell with various properties (such as being in S phase). The main idea is to consider a randomly sampled cell at time t (sampled from all cells that existed until that time, alive or dead) and let χ_A be a characteristic that counts cells that are alive at t and have some property A . Let χ be a characteristic that counts cells that are alive. At time t , the conditional probability that the randomly sampled cell has property A is then

$$P(A|\mathcal{F}_t) = \frac{Z_t^{\chi_A}}{Z_t^\chi}$$

where \mathcal{F}_t is the σ -algebra generated by the reproduction processes of all individuals born up to time t . By (2.4)

$$P(A|\mathcal{F}_t) = \frac{e^{-\alpha t} Z_t^{\chi_A}}{e^{-\alpha t} Z_t^\chi} \rightarrow \frac{E[\widehat{\chi}_A(\alpha)]}{E[\widehat{\chi}(\alpha)]}$$

as $t \rightarrow \infty$. One complication is that the population might go extinct, in which case the probability $P(A|\mathcal{F}_t)$ is not always well-defined. As we deal with cell populations that do not go extinct, we shall not delve deeper into the issue [which also involves possible degeneracy of W associated with the so-called $x \log x$ condition, a fascinating topic in its own right, see Jagers (1975), Lyons et al. (1995), Olofsson (1998, 2009)]. The limit of $P(A|\mathcal{F}_t)$ is called the stable population measure, often denoted $\tilde{P}(A)$. Thus, \tilde{P} takes into account two sources of randomness: the population dynamics and the sampling, the latter being affected by the exponential growth. As the cell populations in Chiorino et al. (2001) are in stable exponential growth, we use the probability measure \tilde{P} in our calculations.

3 The branching process model: theoretical results

In this section we describe the branching process model and state some theoretical results, leading into the data analysis of the next section where explicit assumptions are made regarding phase time parameters.

The data in Chiorino et al. (2001) are obtained by flow cytometry where the last two phases of the cell cycle, G_2 and M , are not distinguishable.

Therefore, we let the cell cycle time be denoted by L where L is the sum of the times of three cell cycle phases: $L = G_1 + S + G_2M$, in the obvious notation. Assume that the lengths G_1, S , and G_2M of the phases are independent continuous random variables (although many of the results are easily rephrased in the case of dependent cell cycle times, using joint distributions rather than products of marginals) and use the notation F_X and f_X for the cdf and pdf of a random variable X . The Malthusian parameter α is determined by the relation

$$2\widehat{F}_L(\alpha) = 1$$

and to count cells in S phase we use the random characteristic

$$\chi_S(a) = I\{G_1 \leq a \leq G_1 + S\} \quad (3.1)$$

which is 1 if the cell is in S phase at age a and 0 otherwise so that $Z_t^{\chi_S}$ gives the number of cells in S phase at time t . The expected value that is needed to compute the limit in (2.4) is

$$E[\chi_S(a)] = P(G_1 \leq a \leq G_1 + S) \quad (3.2)$$

which we can compute once we have explicit distributional assumptions about G_1 and S . The total number of cells alive is counted by the characteristic

$$\chi_L(a) = I\{L \geq a\}$$

which has expected value

$$E[\chi_L(a)] = P(L \geq a)$$

Thus, the fraction of cells in S phase at time t is given by

$$Q(t) = \frac{Z_t^{\chi_S}}{Z_t^{\chi_L}}$$

and our first result expresses the limit of $Q(t)$ as $t \rightarrow \infty$ in terms of the cdf of S , the pdf of G_1 , and α .

Proposition 3.1 *Let $Q(t)$ be as above. Then, with probability 1,*

$$Q(t) \rightarrow 2\alpha \int_0^\infty \int_0^t e^{-\alpha t} (1 - F_S(t - u)) f_{G_1}(u) du dt$$

as $t \rightarrow \infty$.

Proof. Appendix A.

We are interested not only in the limit of $Q(t)$ but also in how the limit is approached. From now on, we shall focus on the expected value of $Q(t)$ and use the approximation

$$E[Q(t)] = E\left[\frac{Z_t^{\chi_s}}{Z_t^{\chi_L}}\right] \approx \frac{E[Z_t^{\chi_s}]}{E[Z_t^{\chi_L}]} \quad (3.3)$$

which can be viewed as a 0th order Taylor approximation where we first consider the function $f(x, y) = x/y$ expanded about a point (a, b) , $f(x, y) \approx a/b$, then let $x = Z_t^{\chi_s}$, $y = Z_t^{\chi_L}$, $a = E[Z_t^{\chi_s}]$ and $b = E[Z_t^{\chi_L}]$ and take expected values. We can now use results from Jagers and Nerman (1984) to deal directly with expressions of the type $E[Z_t^\chi]$ for $t \geq 0$. A brief summary is presented in Appendix B.

As we observe a population where the ancestor starts in S phase, we are in fact observing a time-shifted branching process. Our time zero of observation is really time τ in the branching process where $\tau = G_1 + T$, T being a random variable that gives the position of the ancestor within S phase. Thus, at time t we are in fact observing $Q(t + \tau)$, from now on denoted by $Q_\tau(t)$, where we wish to compute

$$E[Q_\tau(t)] \approx \frac{E[Z_{\tau+t}^{\chi_s}]}{E[Z_{\tau+t}^{\chi_L}]}$$

This quantity has the same limit as $Q(t)$. To see why, let c denote the almost surely constant limit of $Q(t)$ and note that

$$Q_\tau(t) = \frac{Z_{t+\tau}^{\chi_s}}{Z_{t+\tau}^{\chi_L}} = \frac{e^{-\alpha(t+\tau)} Z_{t+\tau}^{\chi_s}}{e^{-\alpha(t+\tau)} Z_{t+\tau}^{\chi_L}} \rightarrow c$$

almost surely as $t \rightarrow \infty$, by (2.4). Since $|Q_\tau(t)| \leq 1$ for all t , dominated convergence yields $E[Q_\tau(t)] \rightarrow c$ as $t \rightarrow \infty$. Also, the limit of the approximating ratio $E[Z_{\tau+t}^{\chi_s}]/E[Z_{\tau+t}^{\chi_L}]$ equals c since

$$\frac{E[Z_{\tau+t}^{\chi_s}]}{E[Z_{\tau+t}^{\chi_L}]} = \frac{E[e^{-\alpha(t+\tau)} Z_{\tau+t}^{\chi_s}]}{E[e^{-\alpha(t+\tau)} Z_{\tau+t}^{\chi_L}]} \rightarrow c$$

by (2.3).

Before we start dealing with $E[Q_\tau(t)]$, let us first state the pdf of the age of the ancestor. As the cell populations in Chiorino et al. (2001) can

be considered in stable exponential growth, we can get the distribution of τ through the asymptotic theory of branching processes, expressed in terms of the distributions of the cell cycle phases. The following proposition is essentially formula (23) in Macdonald, 1978.

Proposition 3.2 *Consider a cell population in stable exponential growth and let τ be the age of a randomly sampled cell that is in S phase. Then τ has pdf*

$$f_\tau(t) = ce^{-\alpha t} \int_0^t (1 - F_S(t - u)) f_{G_1}(u) du$$

where c is a normalizing constant.

Proof. Appendix A.

As an example, consider the simple case where both G_1 and S have exponential distributions with mean 1. Then

$$\begin{aligned} f_\tau(t) &= ce^{-\alpha t} \int_0^t e^{-(t-u)} e^{-u} du \\ &= cte^{-(1+\alpha)t} \end{aligned}$$

and since f_τ must integrate to 1, we get $c = (1 + \alpha)^2$ and we recognize that τ has a $\Gamma(2, 1 + \alpha)$ distribution (more about this distribution in the next section). Note that although the distribution of the third phase G_2M does not appear explicitly, it has an impact on the value of α by (2.1).

To deal with $E[Q_\tau(t)]$, we single out the ancestor and decompose the χ -counted population by adding the contribution of the ancestor and the contributions of the populations stemming from the offspring of the ancestor, a common trick in branching process analysis, see Jagers and Nerman (1984) for details. For any characteristic χ ,

$$Z_t^\chi = \chi(t) + Z_{t-L}^\chi(1) + Z_{t-L}^\chi(2)$$

where L is the lifetime of the ancestor and $Z_{t-L}^\chi(1)$ and $Z_{t-L}^\chi(2)$ denote the two independent branching processes initiated by the children of the ancestor. Now add the ancestor's age τ to t to obtain the expected value

$$\begin{aligned}
E[Z_{\tau+t}^X] &= E[\chi(\tau+t)] + 2E[Z_{\tau+t-L}^X] \\
&= E[\chi(\tau+t)] + 2E[Z_{t-R}^X]
\end{aligned} \tag{3.4}$$

were $R = L - \tau$, the remaining lifetime of the ancestor at time τ . Let us deal with the two terms in (3.4) separately, starting with the first term for the characteristic χ_S . By (3.2),

$$\begin{aligned}
E[\chi_S(\tau+t)] &= P(G_1 \leq \tau+t \leq G_1+S) \\
&= P(G_1 \leq G_1+T+t \leq G_1+S) \\
&= P(S-T \geq t)
\end{aligned}$$

Hence, if we let $X = S - T$, the remaining time the ancestor spends in S phase after observation time 0, the contribution from the ancestor to the time-shifted population is

$$E[\chi_S(\tau+t)] = P(X \geq t) \tag{3.5}$$

where we can express the distribution of X by again invoking asymptotic results from general branching process theory. Note in particular that at the initial observation time $t = 0$ we have $E[\chi_S(\tau)] = 1$ (and even $\chi_S(\tau) \equiv 1$), thus forcing the ancestor to start in S phase. We give the distribution of X next. This distribution is also given in

Proposition 3.3 *Consider a cell population in stable exponential growth and let X be the remaining time a cell that is in S phase spends in that phase. Then X has pdf*

$$f_X(x) = c \frac{d}{dx} \int_0^\infty \int_0^t e^{-\alpha t} F_S(x+t-u) f_{G_1}(u) du dt$$

where c is a normalizing constant.

Note: If f_S is continuous or if $f_S(x + t - u) \leq g(t, u)$ for some function g with

$$\int_0^\infty \int_0^t e^{-\alpha t} g(t, u) f_{G_1}(u) du dt < \infty$$

then differentiation and integration can be interchanged so that

$$f_X(x) = c \int_0^\infty \int_0^t e^{-\alpha t} f_S(x + t - u) f_{G_1}(u) du dt$$

a form which is recognized as formula (24) in Macdonald, 1978.

Proof. Appendix A.

We will use the gamma distribution in our applications, which means that we get the second expression for $f_X(x)$. Note that if S has an exponential distribution, $S \sim \exp(\lambda)$ for some rate λ , then $f_S(s) = \lambda e^{-\lambda s}$ and we get

$$f_X(x) = e^{-\lambda x} c \lambda \int_0^\infty \int_0^t e^{-\alpha t} e^{-\lambda(t-u)} f_{G_1}(u) du dt$$

and as the argument x only appears in the factor $e^{-\lambda x}$ we conclude that $X \sim \exp(\lambda)$ as well, regardless of the distribution of G_1 . Considering the memoryless property of the exponential distribution and the fact that we condition on the cell being in S phase, this observation should come as no surprise.

Next, let us deal with χ_L . For that purpose, let $R = X + G_2M$, the remaining lifetime of a cell in S phase. Thus, the pdf of R is the convolution

$$f_R(r) = \int_0^r f_X(r - u) f_{G_2M}(u) du \quad (3.6)$$

where f_X is given in Proposition 3.3. Let T and X be as above and note that $L = G_1 + S + G_2M$ and $\tau = G_1 + T$ which gives

$$\begin{aligned} E[\chi_L(\tau + t)] &= P(G_1 + S + G_2M \geq G_1 + T + t) \\ &= P(R \geq t) \end{aligned}$$

For the second term in (3.4), note that, for any characteristic χ ,

$$E[Z_{t-R}^\chi] = \int_0^t E[Z_{t-r}^\chi | R = r] f_R(r) dr$$

where f_R is the pdf of R given in (3.6). Note that, conditioned on $R = r$, the conditional expectation $E[Z_{t-r}^\chi | R = r]$ equals $E[Z_{t-r}^\chi]$ since the process starts over and its future is (conditionally) independent of R . Hence, we can rewrite (3.4) as

$$E[Z_{\tau+t}^\chi] = E[\chi(\tau + t)] + 2 \int_0^t E[Z_{t-r}^\chi] f_R(r) dr \quad (3.7)$$

which gives

$$E[Q_\tau(t)] \approx \frac{P(X \geq t) + 2 \int_0^t E[Z_{t-r}^{\chi_S}] f_R(r) dr}{P(R \geq t) + 2 \int_0^t E[Z_{t-r}^{\chi_L}] f_R(r) dr} \quad (3.8)$$

where $P(X \geq t)$ and $P(R \geq t)$ are computed by invoking Proposition 3.3 and (3.6), respectively, $E[Z_{t-r}^\chi]$ can be computed for χ_S and χ_L by the methods outlined in Appendix B, and $f_R(r)$ is given in (3.6). Also note that $E[Q_\tau(t)]$ and its approximating ratio agree at $t = 0$, where they are both equal to 1, and in the limit as $t \rightarrow \infty$.

Let us finally point out that there is an alternative to considering the remaining time X the ancestor spends in S phase, namely, to consider this time as a fraction of the total time in S phase, $X = US$, where U is a random variable with support $[0, 1]$. This is the approach taken by Milotti et al. (2008) where it is assumed that U is uniform on $[0, 1]$. Our last proposition states the cdf of $U = X/S$. It is very similar to the “elapsed proportion of a phase” in Macdonald, 1975.

Proposition 3.4 *Consider a cell population in stable exponential growth, let X and S be as above and let $U = X/S$. The cdf of U is*

$$F(u) = c \int_0^\infty \int_0^t e^{-\alpha t} \left(F_S\left(\frac{t-v}{1-u}\right) - F_S(t-v) \right) f_{G_1}(v) dv dt$$

for $0 \leq u \leq 1$ where c is a normalizing constant.

Proof. Appendix A.

We shall not further utilize this approach, but let us point out that it is clear from the complicated expression of $F(u)$ that U will typically not be uniform. As an example, consider the situation where G_1, S , and G_2M have exponential distributions with mean 1. Simple calculations show that $\alpha = 2^{1/3} - 1 \approx 0.26$ and the cdf of U is

$$F(u) = \frac{0.26u}{0.33 - 0.07u}, \quad 0 \leq u \leq 1$$

where all numbers are rounded to 2 decimals. Clearly U is not uniform, although it is fairly close, see Figure 1. The data in Chiorino et al. (2001)

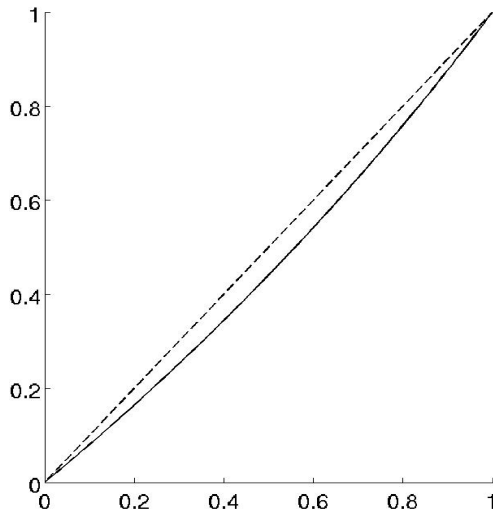


Figure 1: Cdf of U (solid) and of a uniform distribution (dashed).

exhibit an initial linear decline in the fraction of cells in S phase and Milotti et al. (2008) attribute this fact to cells being uniformly positioned within S phase but as we have seen, this is likely not true. An initial linear decline can however be explained without invoking the uniform distribution. An intuitive explanation is that, early on, the population is dominated by the contribution of the ancestor and by (3.5), this contribution equals $P(X \geq t)$. For many distributions, the survival function $P(X \geq t)$ is approximately

linear in the beginning. For example, in the simple case when $S \sim \exp(\lambda)$, we saw that $X \sim \exp(\lambda)$ and hence

$$\begin{aligned} E[Q_\tau(t)] &\approx P(X \geq t) \\ &= e^{-\lambda t} \approx 1 - \lambda t \end{aligned}$$

for small λt . We will elaborate further on this observation in the Discussion section.

4 The branching process model: data analysis

We now assume that the cell cycle times G_1 , S , and G_2M have gamma distributions: $G_1 \sim \Gamma(a_1, b_1)$, $S \sim \Gamma(a_2, b_2)$, and $G_2M \sim \Gamma(a_3, b_3)$. The gamma distribution is a flexible two-parameter family that is commonly used to model lifetimes [Oprea and Kepler (2001), Larsson et al. (2008)]. Specifically, if the parameters are a and b , the probability density function is

$$f(t) = e^{-bt} b^a \frac{t^{a-1}}{\Gamma(a)}, \quad t \geq 0$$

where $\Gamma(a)$ is the gamma function, the mean is a/b , and the variance is a/b^2 . The Malthusian parameter is given by solving the equation $2\hat{F}_L(\alpha) = 1$. By independence we get

$$\hat{F}_L(\alpha) = \hat{F}_{G_1}(\alpha) \hat{F}_S(\alpha) \hat{F}_{G_2M}(\alpha)$$

The Laplace transform for the $\Gamma(a, b)$ distribution is

$$\hat{F}(\alpha) = \int_0^\infty e^{-\alpha t} f(t) dt = \frac{b^a}{(\alpha + b)^a}$$

and hence α is the solution to the equation

$$\frac{2b_1^{a_1} b_2^{a_2} b_3^{a_3}}{(\alpha + b_1)^{a_1} (\alpha + b_2)^{a_2} (\alpha + b_3)^{a_3}} = 1 \tag{4.1}$$

Further on, for our computations we need not only the Laplace transform but the distribution of the total cell cycle time $L = G_1 + S + G_2M$. The phase times have gamma distributions but unless all the b_j are equal, L does not have a gamma distribution. It is possible to express its pdf in closed form as an infinite series [Moschopoulos (1985)] but in order to simplify the computations, we approximate the distribution of L by a gamma distribution. To match the mean and variance of the sum $G_1 + S + G_2M$, this gamma distribution must have parameters a_L and b_L that satisfy

$$\begin{cases} E[L] = \frac{a_L}{b_L} = \frac{a_1}{b_1} + \frac{a_2}{b_2} + \frac{a_3}{b_3} \\ \text{Var}[L] = \frac{a_L}{b_L^2} = \frac{a_1}{b_1^2} + \frac{a_2}{b_2^2} + \frac{a_3}{b_3^2} \end{cases}$$

With our gamma approximation, α is instead the solution to the equation

$$\frac{2b_L^{a_L}}{(\alpha + b_L)^{a_L}} = 1$$

which gives

$$\alpha = b_L(2^{1/a_L} - 1)$$

For one example, let us use a data set from the cell line IGROV1 (ovarian carcinoma) where the estimated parameter values are

$$a_1 = 4, b_1 = 0.44, a_2 = 100, b_2 = 11.5, a_3 = 100, b_3 = 32.2$$

[Lupi et al. (2006) and Paolo Ubezio (personal communication)] which gives $a_L = 20.2$, $b_L = 0.97$, and $\alpha = 0.034$. Inserting this value of α and the a_j and b_j into (4.1) gives the result 0.9997, close enough to 1 to deem our approximation reasonable. It is also easy to verify the practical validity of the approximation by comparing simulated data sets.

Figure 2 shows $Q(t)$ for the IGROV1 parameters given above and for t from 0 to 60 hours. The limit of $Q(t)$ by Proposition 3.1 equals 0.38 (dashed line) in agreement with the value obtained by Chiorino et al. Note the oscillatory pattern which is typical for quantities relating to the cell cycle and shows up in data as well as in models, see for example Bronk et al. (1968), Macdonald (1970), Jagers (1975), Chiorino et al. (2001), Milotti et al (2008), and Olofsson (2008). For details on how to compute $Q(t)$, see Appendix B.

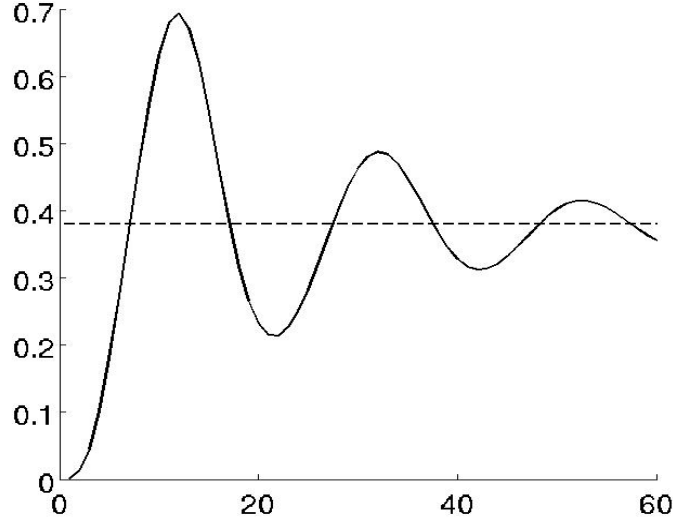


Figure 2: The fraction of cells in S phase converging to 0.38.

In the computation we started at time $t = 0$ which is why the graph in Figure 2 starts at $Q(0) = 0$ (the initial cell starts in G_1 phase). Recall that the data in Chiorino et al. has all cells starting in S phase so we need to consider $E[Q_\tau(t)]$ which is computed according to (3.8). Figure 3 shows $E[Q_\tau(t)]$ for t from 0 to 60 hours (solid line) using the IGROV1 parameter values from above. The figure also displays data from Chiorino et al. (2001) of S phase fractions in an IGROV1 cell line. We have not done any parameter estimation or fitting and it is remarkable how well our model describes the data.

Another quantity of interest is the period between consecutive maxima in the desynchronization curve. As our expression for $E[Q(t)]$ does not have a simple analytic form, we cannot directly establish an expression for the period. However, we can offer the following argument. If cell cycle times were deterministic, the period would equal the cell cycle time. In reality, cell cycle times are random and one might guess that the period instead equals the expected cell cycle time. However, this is not the case because of effects from the exponential growth. Recall the probability measure \tilde{P} from Section 2; it is with respect to this measure we need to compute the expected cell

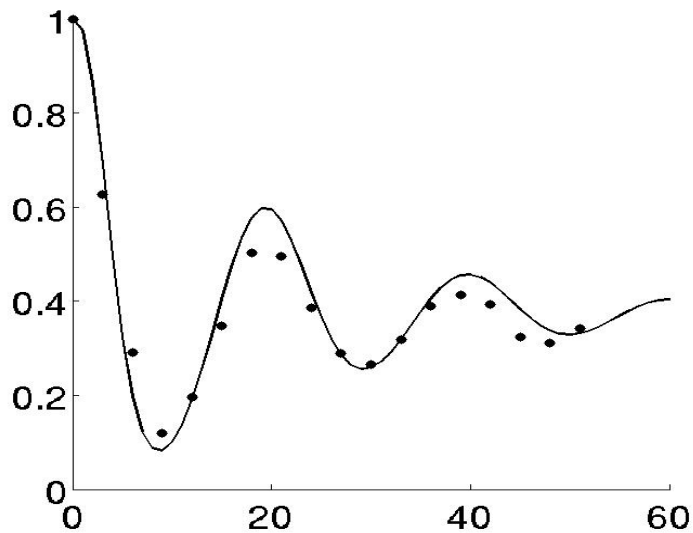


Figure 3: The fraction of cells in S phase—model and data.

cycle time. It turns out that this expected value is the number β defined in (2.2) which is thus a reasonable candidate for the period. In our model where the lifetime L has a $\Gamma(a, b)$ distribution we can find β explicitly as

$$\begin{aligned}\beta &= 2 \int_0^\infty t e^{-at} e^{-bt} b^a \frac{t^{a-1}}{\Gamma(a)} dt \\ &= \frac{a}{b2^{1/a}} = \frac{m}{2^{1/a}}\end{aligned}$$

where $m = E[L] = a/b$ and we recall $\alpha = b(2^{1/a} - 1)$. Chiorino et al. (2001) denotes the period by T and gives the approximation

$$\begin{aligned}T &\approx \frac{m}{1 + \log 2 (\sigma/m)^2} \\ &= \frac{m}{1 + \frac{\log 2}{a}}\end{aligned}$$

since $\sigma/m = 1/\sqrt{a}$ in the $\Gamma(a, b)$ distribution. Now note the first-order Taylor approximation

$$2^x \approx 1 + x \log 2$$

about $x = 0$ to conclude that $\beta \approx T$ unless a is very small. In other words, the approximation in Chiorino et al. is good for the gamma distribution unless the coefficient of variation $\sigma/m = 1/\sqrt{a}$ is large. Calculations indicate that the error is negligible for any realistic values of a , an observation that provides a nice agreement between our branching process model and the deterministic model of Chiorino et al.

Note that the period is always shorter than the mean cell cycle time which makes intuitive sense because at the time of sampling, ancestral lines with many reproduction events are overrepresented, a typical effect of the sampling bias that arises from exponential growth. In contrast, Milotti et al. (2008) present a formula suggesting instead that the period is longer than the mean cell cycle time. Their formula is obtained through analysis of the spectral density of the desynchronization curves but as we have already pointed out, they explicitly disregard exponential growth and hence also the sampling bias that arises from it.

5 Discussion

We have proposed a general branching process model to describe the fraction of cells in S phase in an exponentially growing cell population. The lengths of the cell cycle phases are modeled by gamma distributions whose parameters are taken from previously published data. Given these parameters, we can compute the asymptotic stable fraction of cells in S phase and also investigate how this limit is approached by computing the expected fraction of cells in S phase for any time t . Our model gives very good agreement with published data, showing that the branching process model is indeed a realistic description of how such cell populations evolve.

Our curve shows the typical oscillatory pattern found in the data. It is also noteworthy how the initial “linear” phase arises automatically and, as pointed out in the previous section, is not really linear but an artifact of the distribution of the remaining time in S phase, X (whose distribution is given in Proposition 3.3). Milotti et al. (2008) uses an explicit linear form for the initial time period followed by a shifted damped oscillation, which is why

their model gives a sharp edge where the linear part ends and the oscillation begins. In contrast, our curve is smooth which gives a better description of the data. It should also be noted that our description of the initial time shift as random rather than constant is more accurate. Indeed, the integration in (3.8) affects both period and amplitude which is why the graphs in Figures 2 and 3 are not merely shifted versions of one another with an initial linear part added.

Our purpose was to develop a model that could accurately predict data from desynchronization experiments such as those of Chiorino et al. (2001). A natural continuation for future research would be to develop estimation and curve fitting procedures based on our model and applied to desynchronization data sets. It would also be of interest to investigate the accuracy of the approximation in (3.3) which might involve continuous-time versions of the results in Olofsson and Shaw (2002).

6 Appendix A

In this section we provide proofs of Propositions 3.1–3.4. The proofs are based on Theorems 3.4 and 5.10 from Jagers and Nerman (1984) and for the reader's convenience, we restate them here, together with a result for the total population size Y_t that follows from Lemma 5.2.1 in Jagers (1975).

Theorem 6.1 (Jagers and Nerman (1984)) *Suppose that the mean reproduction process $\mu(dt)$ is nonlattice (cannot be supported by any lattice $\{a, 2a, 3a, \dots\}$), and let $\alpha > 0$ be the solution to the equation $\hat{\mu}(\alpha) = 1$, define $\beta = \int_0^\infty te^{-\alpha t} \mu(dt)$, and let χ be a random characteristic. Assume the following:*

$$(i) \mu(0) < 1 \text{ and } \mu(t) < \infty \text{ for all } t \geq 0$$

$$(ii) \beta < \infty$$

$$(iii) \hat{\mu}(r) < \infty \text{ for some } r < \alpha$$

$$(iv) E \left[\sup_a e^{-\alpha a} \chi(a) \right] < \infty$$

$$(v) \sum_{k=0}^{\infty} \sup_{k \leq a \leq k+1} e^{-\alpha a} E[\chi(a)] < \infty$$

Then

$$E[e^{-\alpha t} Z_t^\chi] \rightarrow \frac{E[\hat{\chi}(\alpha)]}{\beta}$$

as $t \rightarrow \infty$. Next, let Y_t be the total population size (all individuals born, dead or alive) at time t . Then $Y_t < \infty$ almost surely for all $t \geq 0$ and

$$e^{-\alpha t} Z_t^\chi \rightarrow \frac{E[\hat{\chi}(\alpha)]}{\beta} W$$

almost surely on the set of nonextinction $\{Y_t \rightarrow \infty\}$ as $t \rightarrow \infty$, where W is a random variable with mean 1. Further,

$$\frac{Z_t^\chi}{Y_t} \rightarrow E[\hat{\chi}(\alpha)]$$

almost surely on the set of nonextinction $\{Y_t \rightarrow \infty\}$ as $t \rightarrow \infty$.

It follows immediately that if χ_1 and χ_2 are two random characteristics satisfying the conditions of the theorem, we get

$$\frac{E[Z_t^{\chi_1}]}{E[Z_t^{\chi_2}]} \rightarrow \frac{E[\hat{\chi}_1(\alpha)]}{E[\hat{\chi}_2(\alpha)]}$$

as $t \rightarrow \infty$, and also

$$\frac{Z_t^{\chi_1}}{Z_t^{\chi_2}} \rightarrow \frac{E[\hat{\chi}_1(\alpha)]}{E[\hat{\chi}_2(\alpha)]}$$

almost surely on the set of nonextinction as $t \rightarrow \infty$.

Proof of Proposition 3.1. Recalling that we have

$$\mu(dt) = 2F_L(dt) = 2f_L(t)dt$$

it is easily seen that (i)–(iii) hold if L has finite mean (and recall that we use the gamma distribution for the data analysis). Moreover, the characteristic counting all cells is

$$\chi_L(a) = I\{L > a\}$$

which implies that $\chi_L(a) \leq 1$, and (iii) and (iv) follow immediately. Similarly, (iii) and (iv) hold for the characteristic χ_S . Use (3.2) and condition on G_1 to obtain

$$\begin{aligned} E[\widehat{\chi}_S(\alpha)] &= \int_0^\infty e^{-\alpha t} P(G_1 \leq t \leq G_1 + S) dt \\ &= \int_0^\infty e^{-\alpha t} \int_0^t P(u \leq t \leq u + S) f_{G_1}(u) du dt \\ &= \int_0^\infty \int_0^t e^{-\alpha t} (1 - F_S(t - u)) f_{G_1}(u) du dt \end{aligned}$$

and, invoking integration by parts,

$$\begin{aligned} E[\widehat{\chi}_L(\alpha)] &= \int_0^\infty e^{-\alpha t} (1 - F_L(t)) dt \\ &= \left[-\frac{1}{\alpha} e^{-\alpha t} (1 - F_L(t)) \right]_0^\infty - \frac{1}{\alpha} \int_0^\infty e^{-\alpha t} f_L(t) dt \\ &= \frac{1}{2\alpha} \end{aligned}$$

since

$$\int_0^\infty e^{-\alpha t} f_L(t) dt = \frac{1}{2}$$

by the definition of α , and Proposition 3.1 follows. ■

Proof of Proposition 3.2. As previously, let $\chi_S(s) = I\{G_1 \leq s \leq G_1 + S\}$, the indicator that the cell is in S phase at age s . Next, let

$$\chi_a(s) = \chi_S(s) I\{s \leq a\}$$

the indicator that the cell is in S phase and younger than a at age s . The conditional probability that a randomly sampled cell is younger than a given that it is in S phase at time t is then

$$P(\tau \leq a | \text{S phase at time } t) = \frac{Z_t^{\chi_a}}{Z_t^{\chi_S}}$$

which has limit

$$\lim_{t \rightarrow \infty} P(\tau \leq a | \text{S phase at time } t) = \frac{E[\widehat{\chi}_a(\alpha)]}{E[\widehat{\chi}_S(\alpha)]}$$

as $t \rightarrow \infty$. The denominator is the reciprocal of the constant c and is given in the proof of Proposition 3.1 above. The numerator equals

$$E[\widehat{\chi}_a(\alpha)] = \int_0^a \int_0^t e^{-\alpha t} (1 - F_S(t - u)) f_{G_1}(u) du dt$$

and Proposition 3.2 follows. ■

Proof of Proposition 3.3. This proof goes along the same line as the previous proofs, the only difference being that we now need to consider the characteristic

$$\chi_x(s) = I\{G_1 \leq s \leq G_1 + S \leq s + x\}$$

which equals 1 if the cell is in S phase at age s and remains there for at most x more time units. At time t , we have the conditional probability

$$P(X \leq x | \text{S phase at time } t) = \frac{Z_t^{\chi_x}}{Z_t^{\chi_S}}$$

and for its limit as $t \rightarrow \infty$, we let $c = 1/E[\widehat{\chi}_S(\alpha)]$, given above. Further,

$$E[\widehat{\chi}_x(\alpha)] = \int_0^\infty e^{-\alpha t} \int_0^t P(S \leq t + x - u) f_{G_1}(u) du dt$$

and Proposition 3.3 follows. To address the note after the proposition, differentiation under the integral sign is allowed if f_S is continuous, by Leibniz integral rule, or if there is a function g such that $f_S(x + t - u) \leq g(t, u)$ and

$$\int_0^\infty \int_0^t e^{-\alpha t} g(t, u) f_{G_1}(u) du dt < \infty$$

by Theorem (2.27) in Folland (1984). ■

Proof of Proposition 3.4. The characteristic we now use is

$$\chi_u(t) = I\{G_1 \leq t \leq G_1 + S, G_1 + S - t \leq uS\}$$

which has expected value

$$E[\chi_u(t)] = P\left(G_1 \leq t \leq G_1 + S, S \leq \frac{t - G_1}{1 - u}\right)$$

and as the proof follows the pattern of the previous proofs, we leave out the details. \blacksquare

7 Appendix B

In Section 4, we used the approximation

$$E\left[\frac{Z_t^{\chi S}}{Z_t^{\chi L}}\right] \approx \frac{E[Z_t^{\chi S}]}{E[Z_t^{\chi L}]}$$

where $E[Z_t^{\chi S}]$ and $E[Z_t^{\chi L}]$ were computed numerically. To do so, standard results from renewal theory yield

$$E[Z_t^\chi] = E[\chi] * \nu(t) = \int_0^t E[\chi(t - u)]\nu(du)$$

where ν is the renewal measure

$$\nu(du) = \sum_{n=0}^{\infty} \mu^{*n}(du)$$

μ^{*n} being the n -fold convolution of μ (where μ^{*0} by definition equals δ_0 , the unit point mass at 0). If $\mu(0) < 1$ and $\mu(t) < \infty$ for all $t \geq 0$, it is known that $\nu(t) < \infty$ for all $t \geq 0$, see Lemma 5.2.1 in Jagers (1975). In our cell population reproduction $\mu(du) = 2F(du)$ so that

$$\mu^{*n}(du) = 2^n F^{*n}(du)$$

The expression is intuitively reasonable. There are 2^n individuals in the n th generation and F^{*n} is the distribution function of the sum of n independent lifetimes. Hence, an individual in the n th generation is born before time t with probability $F^{*n}(t)$, so the expected number of individuals from the n th generation that are born before t equals $2^n F^{*n}(t)$. Summing over n gives the expected number of individuals born before time t which is indeed the interpretation of the renewal measure ν .

With our assumption that $L \sim \Gamma(a, b)$, we get, by additivity of the gamma distribution,

$$\mu^{*n}(du) = 2^n f^{*n}(u) du = 2^n e^{-bt} b^{na} \frac{u^{na-1}}{\Gamma(na)} du$$

and we compute expressions of the type

$$E[Z_t^\chi] = E[\chi(t)] + \sum_{n=1}^{\infty} 2^n \int_0^t E[\chi(t-u)] f^{*n}(u) du$$

where $E[\chi(t)]$ is the term for $n = 0$. The functions that are to be integrated are $E[\chi_S(t)]$ and $E[\chi_L(t)]$. By (3.2) we get

$$\begin{aligned} E[\chi_S(t)] &= P(G_1 \leq t \leq G_1 + S) \\ &= \int_0^t P(G_1 \leq t \leq G_1 + S | G_1 = v) f_{G_1}(v) dv \\ &= \int_0^t P(S > t - v) f_{G_1}(v) dv \end{aligned}$$

and

$$E[\chi_L(t)] = P(L > t)$$

We can now compute $E[Z_t^{\chi_S}]$ and $E[Z_t^{\chi_L}]$ for any t .

8 Acknowledgments

The authors thank Paolo Ubezio at the Mario Negri Institute for Pharmaceutical Research in Milano for generously sharing and patiently explaining data sets, and Mark Embree and Jeff Hokanson at Rice University for invaluable help with numerical integration. We also thank an anonymous referee for alerting us to the publications by P.D.M. Macdonald.

9 References

O. Arino, E.Sánchez, G. Webb, Necessary and sufficient conditions for asynchronous exponential growth in age structured cell populations with quiescence, *J. Math. Appl.* 215 (1997), 499.

- B. V. Bronk, G. J. Dienes, A. Pasking, The stochastic theory of cell proliferation, *Biophys.J.*, 8, (1968), 1353.
- B. V. Bronk (1969) On radioactive labeling of proliferating cells: the graph of labeled mitoses. *J. Theoret. Biol.*, 22, (1969), 468.
- G. Chiorino, J. A. J. Metz, D. Tomasoni, P. Ubezio, Desynchronization rate in cell populations: mathematical modeling and experimental data., *J. Theor. Biol.* 208 (2001), 185.
- J. Dyson, R. Villella-Bressan, G. Webb, Asynchronous exponential growth in an age structured population of proliferating and quiescent cells, *math, Biosci.* 177 & 178 (2002), 73.
- G.B. Folland, *Real Analysis* (1st edition), Wiley, New York (1984).
- M. Gyllenberg, W. Webb, Age-size structure in populations with quiescence, *Math. Biosci.* 86 (1987), 67.
- P. Jagers, *Branching Processes with Biological Applications*, Wiley, London, (1975).
- P. Jagers, O. Nerman, The growth and composition of branching populations, *Adv. Appl. Prob.* 16 (1984), 221.
- S. Larsson, T. Ryden, U. Holst, S. Oredsson and M. Johansson, Estimating the total rate of DNA replication using branching processes. *Bulletin of Mathematical Biology* 70(8) (2008), 2177.
- M. Lupi, P. Capella, G. Matera, C. Natoli, P. Ubezio, Interpreting cell cycle effects of drugs: The case of melphalan. *Cancer Chemother Pharmacol* 57(4) (2006), 443.
- R. Lyons, R. Pemantle, Y. Peres, Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Annals of Probability* 3 (1995), 1125.
- P.D.M. Macdonald, Statistical inference from the fraction labelled mitoses curve. *Biometrika*, 57 (1970), 489.

- P.D.M. Macdonald, The distribution of DNA in exponentially-growing cell populations. In *Perspectives in Probability and Statistics: Papers in honour of M.S. Bartlett on the occasion of his sixty-fifth birthday.* (Ed. J. Gani), Applied Probability Trust & Academic Press, (1974), 387.
- P.D.M. Macdonald, Age distributions in the general cell kinetic model. In *Biomathematics and Cell Kinetics*, (Ed., A.-J. Valleron and P.D.M. Macdonald), Elsevier/North-Holland Biomedical Press, (1978), 3.
- P.D.M. Macdonald, Towards an exact analysis of stathmokinetic and continuous-labelling experiments. In *Biomathematics and Cell Kinetics*, (Ed. M. Rotenberg), Elsevier/North-Holland Biomedical Press, (1981), **PAGE**.
- P.G. Moschopoulos, The distribution of the sum of independent gamma random variables, *Ann. Inst. Statist. Math* 37 (1985), 541.
- E. Milotti, A. Del Fabbro, C. Dalla Pellegrina, R. Chignola, Statistical approach to the analysis of cell desynchronization data, *Physica A* 387 (2008), 4204.
- P. Olofsson, The $x \log x$ condition for general branching processes. *J. Appl. Prob.* 35 (1998), 537.
- P. Olofsson, A stochastic model of a cell population with quiescence. *Journal of Biological Dynamics* 2(4) (2008), 386.
- P. Olofsson, Size-biased branching population measures and the multi-type $x \log x$ condition. *Bernoulli* (2009), to appear.
- P. Olofsson and C.A. Shaw, Exact sampling formulas for multi-type Galton-Watson processes. *Journal of Mathematical Biology* 45 (2002), 279.
- M. Oprea and T.B. Kepler, Improved inference of mutation rates: II. Generalization of the Luria-Delbruck distribution for realistic cell-cycle time distributions. *Theoretical Population Biology* 59(1) (2001), 49.