

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

Tibor Antal¹ and P L Krapivsky²

¹ Program for Evolutionary Dynamics, Harvard University, Cambridge, MA 02138, USA

² Department of Physics, Boston University, Boston, MA 02215, USA
E-mail: tibor_antal@harvard.edu and paulk@bu.edu

Received 13 June 2010

Accepted 6 July 2010

Published 30 July 2010

Online at stacks.iop.org/JSTAT/2010/P07028

doi:10.1088/1742-5468/2010/07/P07028

Abstract. We study a two-type branching process which provides an excellent description of experimental data on cell dynamics in skin tissue. The model involves only a single type of progenitor cell, that is it does not require support from a self-renewed population of stem cells. The progenitor cells divide and may differentiate into post-mitotic cells. We derive an exact solution for the generating function of the number of cells of different types. The generating function is then used to numerically determine the time dependent probability distribution of the numbers of the two types of cells. We also deduce large time asymptotic behaviors drawing on exact results and on a diffusion approximation.

Keywords: exact results, stochastic processes (theory), population dynamics (theory)

ArXiv ePrint: [0908.0484](http://arxiv.org/abs/0908.0484)

Contents

1. Introduction	2
2. Model	4
3. Generating function	5
4. Elementary solutions	7
5. Explicit results at a special point	8
6. General results	10
7. Scaling limit	13
8. Fokker–Planck approximation	16
9. Discussion	18
Acknowledgments	19
Appendix: Special cases	19
A.1. Horizontal line $r = 1/4$	19
A.2. Special curve $\gamma = 2r/(1 - 2r)$	20
References	21

1. Introduction

Understanding the kinetics (homeostasis) of cells in adult mammalian tissues has long been a major challenge in biology. Recent progress in experimental methods made it feasible to label individual cells *in vivo*, and follow their fate and that of their progeny [2, 3]. This powerful genetic labeling technique has enabled *in vivo* experiments in the outmost layer of skin (epidermis) of the tail in adult mice [1]. Individual cells in the basal layer of the epidermis have been marked by a fluorescent genetic label, and the size of the clone (all living progenies of a cell) of each single marked cell has been measured at different times. This has provided the data on the evolution of the clone size distribution in the basal layer of the epidermis.

The prevailing model of epidermal homeostasis has involved long-lived stem cells generating short-lived populations of transit-amplifying (TA) cells that differentiate into non-proliferating (post-mitotic) cells [4, 5]. The stem-TA hypothesis predicts that the clones of TA cells should disappear (after sufficiently long time), while the existing clones should be small and associated with stem cells. Strikingly, the fraction of remaining clones was found to decrease inversely proportional to time; accordingly the average size of existing clones scales linearly with time. This remarkable scaling behavior calls for a totally different model of epidermal homeostasis. Clayton *et al* [1] proposed a model of cell division and differentiation which manifestly obeys the observed scaling behavior and provides an excellent fit to more subtle characteristics. A gratifying property of the model suggested in [1] is that it is *simpler* than the stem-TA model: The new model involves

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

only a single type of committed progenitor cell, and in particular, stem-cell proliferation is not required for epidermal homeostasis.

Thus the model describes the population of cells of two types: proliferating cells (type A) divide and eventually differentiate into non-proliferating cells (type B), which leave the basal layer and migrate to the epidermal surface where they are shed. More precisely, the cell population evolves according to the continuous time, constant rate, two-type branching process

$$\begin{aligned} A \rightarrow AA &\quad \text{at rate } r \\ A \rightarrow AB &\quad \text{at rate } 1 - 2r \\ A \rightarrow BB &\quad \text{at rate } r \\ B \rightarrow \emptyset &\quad \text{at rate } \gamma. \end{aligned} \tag{1}$$

Here we set the overall cell division rate to unity. Note that the model is assumed to be critical, that is the division rates corresponding to the channels $A \rightarrow AA$ and $A \rightarrow BB$ are the same. Due to this symmetry, the average population size of progenitor cells remains constant, as is required by the steady-state assumption. The average population size of the post-mitotic cells is also constant.

This model has been tested in experiment [1]. Since progenitor (A) and post-mitotic cells (B) were experimentally indistinguishable, the clone size distribution of the total cell count was measured at different times from 2 weeks to one year. The authors presented excellent fits to the experimental data by the numerical solution of the model. The division rate was independently determined to be 1.1/week (denoted by λ in [1], and set to unity in (1)). Hence, two parameters were used to fit the experimental data, and their values were found to be $r = 0.08$ and $\gamma = 0.28$ (our γ corresponds to Γ/λ of [1]).

So far the model has been experimentally tested only in mice tail skin. There are still technical constraints preventing the quantitative tests of the model in other tissues, but those problems are temporary. The model challenges the necessity of stem-cell proliferation for the homeostasis of epidermis [6]. There is also growing evidence [7, 8] that stem cells do not contribute to the maintenance of various healthy adult tissues. Hence the two-type branching process (1) may find a broad range of applications and, therefore, it is highly desirable to possess an exact solution. Despite its apparent simplicity, the branching process (1) has not been solved, although some exact and asymptotic behaviors have been found [1, 9, 10]. In this paper we apply generating function techniques to obtain an exact analytic solution, as well as approximate methods to derive asymptotic limits.

Branching processes have been extensively used to model proliferation of differentiating cells, especially in the hemopoietic (blood production) system [11, 12]; see also other references in section 6.9.1 in [13]. An interesting multi-type model has also been proposed recently in [14, 15]. These studies mainly rely on numerical solutions, while analytic treatment is restricted to obtaining average quantities (or second moments).

The rest of the paper is organized as follows. In section 2, we introduce the model and discuss its basic behavior. After presenting the generating function methods in section 3, we provide an elementary solution on a special line in the parameter space in section 4. The model admits a neat explicit solution at the special point $\gamma = 1, r = 1/4$, which is discussed in section 5. Our main result for the generating function with arbitrary parameter values is presented in section 6, where we also outline an efficient numerical method that allows us to obtain the probabilities of having a certain number of cells at

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

a given time. We discuss the large time asymptotic behavior in section 7, and derive additional scaling properties by means of the Fokker–Planck method in section 8. Final remarks are presented in section 9.

2. Model

The model involves two types of cells, A and B . Type A cells (progenitor cells) are able to divide (proliferate) and differentiate, B cells (post-mitotic cells) do not divide, they just disappear (leave the basal layer). More precisely, the two cell populations evolve according to the two-type branching process (1). The probability $P_{m,n}(t)$ of having m copies of A , and n copies of B at time t satisfies

$$\begin{aligned} \frac{dP_{m,n}}{dt} = & r(m-1)P_{m-1,n} + (1-2r)mP_{m,n-1} + r(m+1)P_{m+1,n-2} \\ & + \gamma(n+1)P_{m,n+1} - (m+\gamma n)P_{m,n}. \end{aligned} \quad (2)$$

The consecutive gain terms on the right-hand side of equation (2) merely describe the contributions of the channels (from top to bottom) of the two-type branching process (1). To determine the clone size distribution we start with a single A cell, that is

$$P_{m,n}(t=0) = \delta_{m,1}\delta_{n,0}. \quad (3)$$

We are interested in the full distribution $P_{m,n}(t)$ and also in the reduced probability distribution $\Pi_s(t)$ of having $s = m + n$ total cells at time t ; the latter distribution is directly probed in experiments. Needless to say,

$$\Pi_s(t) = \sum_{m+n=s} P_{m,n}(t). \quad (4)$$

Let us first determine the probability distribution $P_m(t)$ of having m cells of type A . This probability distribution is readily found since B cells do not affect A cells, and A cells alone evolve according to the critical branching process



The solution, for the initial condition $P_m(t=0) = \delta_{m,1}$, is [16]

$$P_m(t) = \begin{cases} \frac{1}{(1+rt)^2} \left(\frac{rt}{1+rt}\right)^{m-1} & \text{for } m \geq 1 \\ \frac{rt}{1+rt} & \text{for } m = 0. \end{cases} \quad (6)$$

Notice that the average number of A cells remains constant,

$$\langle m \rangle = \sum_{m \geq 0} mP_m(t) = 1, \quad (7)$$

throughout the evolution. This is of course a general property of the critical branching process.

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

We can also compute the average number of post-mitotic cells $\langle n \rangle = \sum_{m,n \geq 0} n P_{m,n}(t)$. Indeed, this quantity satisfies an exact rate equation

$$\frac{d\langle n \rangle}{dt} = \langle m \rangle - \gamma \langle n \rangle. \quad (8)$$

The gain term on the right-hand side of (8) follows from the second and third channels (from top to bottom) of the two-type branching process (1); the loss term corresponds to the last channel. Using $\langle m \rangle = 1$ and $\langle n \rangle|_{t=0} = 0$ we solve (8) to yield

$$\langle n \rangle = \frac{1 - e^{-\gamma t}}{\gamma}. \quad (9)$$

Therefore the total average number of cells is given by

$$\langle s \rangle = 1 + \frac{1 - e^{-\gamma t}}{\gamma}. \quad (10)$$

Note that the fraction of type *A* cells (the fraction of the expected values) is asymptotically

$$\rho \equiv \frac{\langle m \rangle}{\langle s \rangle} = \frac{\gamma}{1 + \gamma}. \quad (11)$$

These exact expressions for the average population sizes are useful and e.g. the fraction of type *A* cells (11) will appear in numerous latter formulae. The full description of the clone size requires analyzing an infinite set of master equations (2). We shall perform such analysis using generating function techniques.

3. Generating function

We define the generating function of $P_{m,n}(t)$ as

$$F(x, y, t) = \sum_{m,n=0}^{\infty} x^m y^n P_{m,n}(t). \quad (12)$$

Note that (2) is valid for all $m, n \geq 0$, if we define $P_{m,n} \equiv 0$ for all $m < 0$ or $n < 0$. (Such systems are said to have *natural* boundary conditions [17].) We multiply both sides of (2) by $x^m y^n$ and sum over all values of $m, n \geq 0$. Using the identities $mx^m = x\partial_x x^m, ny^n = y\partial_y y^n$, where $\partial_x = \partial/\partial x, \partial_y = \partial/\partial y$, we arrive at the (forward Kolmogorov) partial differential equation

$$\partial_t F + [x(1 - y) - r(x - y)^2] \partial_x F + \gamma(y - 1) \partial_y F = 0. \quad (13)$$

The initial condition (3) corresponding to a single initial *A* cell becomes

$$F(x, y, t = 0) = x. \quad (14)$$

Thus we need to solve the partial differential equation (13) subject to (14). Mathematically, equation (13) is a hyperbolic partial differential equation and it can be analyzed using the method of characteristics [18]. Apart from the usual forward Kolmogorov equation (13), the probability distribution $P_{m,n}(t)$ also satisfies the equivalent backward Kolmogorov equations. In this paper we employ these latter equations for the

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

only reason that they are technically somewhat easier to tackle in the present case. These backward equations describe the change of probabilities at present by accounting for events happening to the initial cell. For example an initial A cell becomes two A cells at rate r , in which case we need the probability of having (m, n) cells at time t evolved from *two* initial A cells, which we denote by $P_{m,n}^{AA}$. By writing similarly all possible transitions that can happen to an initial A or B cell, we arrive at the rate equations

$$\begin{aligned}\frac{dP_{m,n}^A}{dt} &= rP_{m,n}^{AA} + (1 - 2r)P_{m,n}^{AB} + rP_{m,n}^{BB} - P_{m,n}^A \\ \frac{dP_{m,n}^B}{dt} &= \gamma\delta_{m,0}\delta_{n,0} - \gamma P_{m,n}^B\end{aligned}\quad (15)$$

The last terms of both equations are the loss terms. The gain term in the second equation $\gamma\delta_{m,0}\delta_{n,0}$ represents the case when a B cell dies at rate γ , and hence it has zero offspring at time t . We again multiply both sides of (15) by $x^m y^n$ and sum over all values of $m, n \geq 0$. We also use the independence of the progenies of two initial cells, hence for example the generating function of $P_{m,n}^{AA}$ simply becomes F_A^2 . Now we need two generating functions F_A and F_B , where the subscripts refer to the type of the single initial cell. (For the forward case we only examine the interesting generating function $F \equiv F_A$.) The initial conditions are

$$F_A(x, y, t = 0) = x \quad F_B(x, y, t = 0) = y \quad (16)$$

and the coupled backward Kolmogorov equations read

$$\partial_t F_A = rF_A^2 + (1 - 2r)F_A F_B + rF_B^2 - F_A \quad (17a)$$

$$\partial_t F_B = \gamma(1 - F_B). \quad (17b)$$

These equations can also be written down directly [16]. Indeed, the negative terms $(-F_A, -\gamma F_B)$ describe the disappearance of a cell, and the positive terms stand for the created new cells, with the corresponding rates.

Equation (17b) is immediately solved to give

$$F_B = ye^{-\gamma t} + 1 - e^{-\gamma t} \equiv f. \quad (18)$$

This is not surprising, of course: starting with a single B cell, the system will either contain the initial B cell (this occurs with probability $e^{-\gamma t}$) or no cells at all. Substituting (18) into (17a) and changing the variable from t to f we obtain

$$\gamma(1 - f)\partial_f F = r(F - f)^2 + F(f - 1) \quad (19)$$

where we dropped the subscript A so that $F \equiv F_A$. We further simplify the above equation by changing variable f to $u = 1 - f = (1 - y)e^{-\gamma t}$. The function $F(u)$ then satisfies

$$\gamma u F' = Fu - r(F + u - 1)^2 \quad (20)$$

with initial condition $F(u = 1 - y) = x$. In equation (20) and later the prime denotes the derivative with respect to u . Note that the forward equation (13) leads to the same equation (20) via the method of characteristics.

Equation (20) is an ordinary differential equation of the first order. Yet it is nonlinear and could be unsolvable as it belongs to the family of Riccati equations. Riccati equations

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

are in principle intractable, yet there are two tricks which sometimes allow one to solve certain Riccati equations [19]. One is based on the reduction to the *linear* ordinary differential equation of the second order, the Sturm–Liouville equation. Another trick applies if we manage to find a special solution. We shall see that both tricks lead to success. Let us begin with the more elementary second approach.

4. Elementary solutions

The idea is to guess one solution $F_*(u)$ irrespective of whether it satisfies the initial condition or not. Having found such a special solution, one then seeks a general solution in the form

$$F(u) = F_*(u) + \frac{1}{V(u)}. \quad (21)$$

The function $V(u)$ satisfies a *linear* differential equation which is readily solvable.

The form of (20) suggests seeking a special solution as a polynomial:

$$F_*(u) = A_0 + A_1 u + \cdots + A_p u^p. \quad (22)$$

Here A_0, A_1, \dots, A_p are constants and $A_p \neq 0$, so that the polynomial (22) has degree p . Noting that uF' is the polynomial of degree p , uF is the polynomial of degree $p+1$, and $(F - 1 + u)^2$ is the polynomial of degree $2p$, equating the highest degree in u would be possible only if $p+1 = 2p$, i.e. $p=1$. Thus the polynomial solution should be a linear function of u ,

$$F_*(u) = A_0 + A_1 u. \quad (23)$$

Plugging (23) into (20) we find that the matching is achieved (that is, the ansatz (23) works) if $A_0 = 1, A_1 = 1/\gamma$ and the parameters r, γ are related via

$$r = \frac{\gamma}{(1+\gamma)^2}. \quad (24)$$

The prescription (21) tells us to seek the general solution in the form

$$F = 1 + \frac{u}{\gamma} + \frac{1}{V(u)}. \quad (25)$$

By inserting (25) into (20) we arrive at a linear ordinary differential equation

$$V' + \hat{\gamma}V - \frac{1}{(1+\gamma)^2} \frac{1}{u} = 0, \quad \hat{\gamma} = \frac{1}{\gamma} \frac{\gamma-1}{\gamma+1}. \quad (26)$$

The homogeneous part has solution $e^{-\hat{\gamma}u}$ and therefore the general solution to (26) is sought as $V = e^{-\hat{\gamma}u} W$. The auxiliary function W obeys

$$W' = \frac{1}{(1+\gamma)^2} \frac{e^{\hat{\gamma}u}}{u} \quad (27)$$

which is solved to yield

$$W = \frac{Ei(\hat{\gamma}u)}{(1+\gamma)^2} + \text{const.} \quad (28)$$

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

Here $Ei(x) = -\int_{-x}^{\infty} d\xi e^{-\xi}/\xi$ is the exponential integral. The constant and the choice of the appropriate lower limit in the integral in equation (28) are fixed by the initial condition. Recall that initially we have $F(u = 1 - y) = x$. Hence (25) gives

$$x - 1 + \frac{y - 1}{\gamma} = \frac{1}{V_0} \quad (29)$$

and therefore

$$W_0 = V_0 e^{\hat{\gamma} u_0} = V_0 e^{\hat{\gamma}(1-y)} = \frac{e^{\hat{\gamma}(1-y)}}{x - 1 + (y - 1)/\gamma}. \quad (30)$$

Using (28) and (30) we obtain

$$W = \frac{e^{\hat{\gamma}(1-y)}}{x - 1 + (y - 1)/\gamma} - \mathcal{E}(y, t) \quad (31)$$

where we used the shorthand notation

$$\mathcal{E}(y, t) = \frac{Ei[\hat{\gamma}(1-y)] - Ei[\hat{\gamma}(1-y)e^{-\gamma t}]}{(1 + \gamma)^2}. \quad (32)$$

Combining (25) and (31) we arrive at

$$F(x, y, t) = 1 + \gamma^{-1} e^{-\gamma t} (1 - y) + e^{\hat{\gamma} e^{-\gamma t} (1-y)} \left[\frac{e^{\hat{\gamma}(1-y)}}{x - 1 + (y - 1)/\gamma} - \mathcal{E}(y, t) \right]^{-1}. \quad (33)$$

The exact solution (33) for the generating function can in principle be expanded in x and y to yield the probability distribution $P_{m,n}$ for arbitrary m, n . For instance, the system is empty with probability

$$P_{0,0} = \Pi_0 = 1 + \gamma^{-1} e^{-\gamma t} - \frac{\exp(\hat{\gamma} e^{-\gamma t})}{\rho e^{\hat{\gamma}} + \mathcal{E}(0, t)} \quad (34)$$

with $\rho = \gamma/(1 + \gamma)$, see equation (11). The expressions for the clone size distribution are simple when $n = 0$, that is for the clones without post-mitotic cells. Expanding $F(x, 0, t)$ in powers of x and using $F(x, 0, t) = \sum_{m \geq 0} x^m P_{m,0}(t)$ we obtain

$$P_{m,0} = \frac{\rho \exp(\hat{\gamma} e^{-\gamma t} + \hat{\gamma})}{[\rho e^{\hat{\gamma}} + \mathcal{E}(0, t)]^2} \rho^m \left[\frac{\mathcal{E}(0, t)}{\rho e^{\hat{\gamma}} + \mathcal{E}(0, t)} \right]^{m-1}. \quad (35)$$

The probabilities $P_{m,n}$ quickly become very unwieldy for $n > 0$.

5. Explicit results at a special point

At the special point $\gamma_c = 1, r_c = 1/4$ in the parameter space we can solve everything explicitly. Indeed, in this case $\hat{\gamma} = 0$ and (33) becomes

$$F(x, y, t) = 1 + (1 - y) e^{-t} + \frac{1}{(x + y - 2)^{-1} - t/4}. \quad (36)$$

Let us first extract the reduced distribution. Writing $x = z, y = z$ and noting that

$$G(z, t) \equiv F(z, z, t) = \sum_{s \geq 0} z^s \Pi_s(t) \quad (37)$$

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

we conclude that

$$G(z, t) = 1 + (1 - z) e^{-t} - \frac{4}{t} + \frac{4}{t} \left[1 + \frac{t}{2} - \frac{t}{2} z \right]^{-1}. \quad (38)$$

Expanding the latter expression in z around $z = 0$ we get

$$\Pi_0 = 1 + e^{-t} - \frac{4}{t+2} \quad (39a)$$

$$\Pi_1 = -e^{-t} + \frac{8}{(t+2)^2} \quad (39b)$$

$$\Pi_s = \frac{8}{(t+2)^2} \left[\frac{t}{t+2} \right]^{s-1}, \quad s \geq 2. \quad (39c)$$

In the scaling region

$$s \rightarrow \infty, \quad t \rightarrow \infty, \quad \frac{s}{t} = \text{finite}, \quad (40)$$

equation (39c) acquires a scaling form

$$\Pi_s \simeq \frac{8}{t^2} e^{-2s/t}. \quad (41)$$

Recall that the exact expression (6) for the distribution of A cells also acquires an asymptotic scaling form; in the present case $r = r_c = 1/4$ it is given by

$$P_m(t) = \frac{16}{(t+4)^2} \left[\frac{t}{t+4} \right]^{m-1} \simeq \frac{16}{t^2} e^{-4m/t}. \quad (42)$$

Generally, by expanding (36), we obtain

$$P_{0,1} = -e^{-t} + \frac{4}{(t+2)^2} \quad (43)$$

and, for $(m, n) \neq (0, 0), (0, 1)$,

$$P_{m,n} = \frac{4}{(t+2)^2} \left[\frac{t}{2(t+2)} \right]^{m+n-1} \binom{m+n}{m}. \quad (44)$$

The probability that the system is empty is $P_{0,0} = \Pi_0$, so it is given by equation (39a).

The clone size distribution greatly simplifies at this special point due to a mapping of our two-type branching process onto a single-type critical branching process. Indeed, at $\gamma = 1, r = 1/4$, the process can be reformulated as



where we assign the type A or B to each cell independently with probability $1/2$. This mapping holds if also initially we have an A or a B cell equiprobably. If the initial cell is type A , then from the solution for a single initial B cell (18), and from the solution (6) of (45), we recover the behavior (43) and (44) due to the linearity of the problem.

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

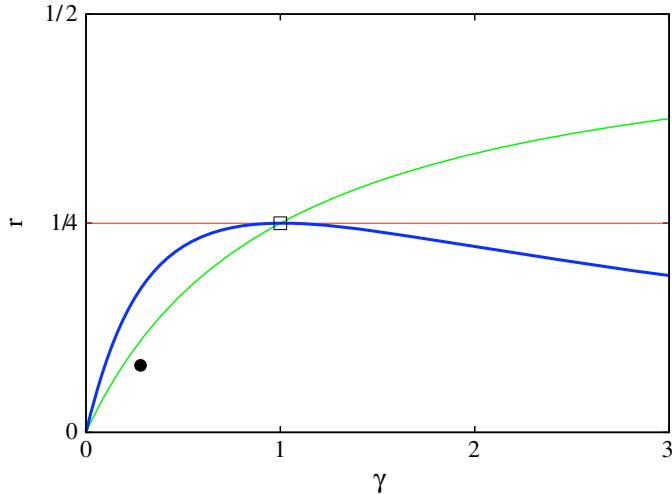


Figure 1. The permissible range in the model parameter space is $0 < r \leq 1/2$ and $0 < \gamma < \infty$. Shown are the curves [$r = 1/4$, $r = \gamma/(1 + \gamma)^2$ thick blue, and $\gamma = 2r/(1 - 2r)$ thin green] where the solution has somewhat simpler forms. At the intersection of these three curves ($\gamma = 1, r = 1/4$, \square) the solution is particularly neat (see section 5). The experimentally measured [1] parameter values ($\gamma = 0.28, r = 0.08$) for mice epidermis are depicted by the \bullet symbol.

6. General results

In section 4 we have found an explicit, exact expression for the generating function, equation (33), which is valid on the curve (24). This curve misses the parameter values ($\gamma = 0.28, r = 0.08$) experimentally measured in mice tail epidermis [1], see figure 1. In different tissues the parameters will probably take different values, so it is desirable to possess a solution in the whole range of parameters, i.e. in the strip $0 < r \leq 1/2$ and $0 < \gamma < \infty$. Fortunately, the reduction of the Riccati equation to the Sturm–Liouville equation leads to a known equation.

We start with the general backward equation (20) which we re-write in a canonical form

$$F' = AF^2 + BF + C. \quad (46)$$

The coefficients of the quadratic polynomial on the right-hand side are

$$A = -\frac{r}{\gamma u}, \quad B = \frac{1}{\gamma} \left[1 + 2r \frac{1-u}{u} \right], \quad C = -\frac{r}{\gamma u} (1-u)^2. \quad (47)$$

To transform the Riccati equation (46) into the Sturm–Liouville equation we perform the standard procedure [19], namely we write $F(u)$ as

$$F = -\frac{z'}{Az} = -\frac{(\log z)'}{A}. \quad (48)$$

After this transformation, the first order nonlinear equation (46) turns into a second order linear differential equation

$$z'' + \alpha z' + \beta z = 0 \quad (49)$$

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

with

$$\begin{aligned}\alpha &= -\left(\frac{A'}{A} + B\right) = \frac{\gamma - u - 2r(1-u)}{u\gamma} \\ \beta &= AC = \left[\frac{r(1-u)}{u\gamma}\right]^2.\end{aligned}\quad (50)$$

Now in (49) the first derivative can be canceled by writing $z = \Phi Z$, with the condition $\Phi' = -\alpha\Phi/2$, which leads to

$$\Phi = e^{-(1/2) \int^u \alpha(u') du'}. \quad (51)$$

Then (49) becomes a Schrödinger equation for $Z(u)$

$$Z'' + \left(\frac{4r-1}{4\gamma^2} + \frac{\gamma(1-2r)-2r}{2u\gamma^2} + \frac{1}{4u^2}\right) Z = 0. \quad (52)$$

Equation (52) resembles the Whittaker equation. Re-scaling the variable u and making changes in notations

$$g = \frac{uv}{\gamma}, \quad v = \sqrt{1-4r}, \quad w = \frac{\gamma(1-2r)-2r}{2\gamma v} \quad (53)$$

we indeed recast equation (52) into a canonical Whittaker differential equation

$$\frac{d^2Z}{dg^2} + \left(-\frac{1}{4} + \frac{w}{g} + \frac{1}{4g^2}\right) Z = 0. \quad (54)$$

Its solution, up to an irrelevant constant factor, is

$$Z(g) = M_{w,0}(g) + CW_{w,0}(g) \quad (55)$$

where M and W are the Whittaker functions [20], and C is a constant to be determined from the boundary conditions.

Now we have to re-express the solution of equation (55) in terms of the original variables. Following the steps that have been made, but backwards, we obtain

$$F(u) = -\frac{[\log z(u)]'}{A} = \frac{\gamma u}{r} [\log Z(u) + \log \Phi(u)']. \quad (56)$$

Using $[\log \Phi(u)]' = -\alpha/2$, see equation (51), we get

$$F(u) = \frac{uv}{r} \cdot \frac{\partial_g M_{w,0}(g) + C \partial_g W_{w,0}(g)}{M_{w,0}(g) + CW_{w,0}(g)} + 1 - u + \frac{u - \gamma}{2r}. \quad (57)$$

Noting that

$$\begin{aligned}\partial_g M_{w,0}(g) &= \frac{(g-2w)M_{w,0}(g) + (1+2w)M_{1+w,0}(g)}{2g} \\ \partial_g W_{w,0}(g) &= \frac{(g-2w)W_{w,0}(g) - 2W_{1+w,0}(g)}{2g}\end{aligned}\quad (58)$$

we simplify (57) and arrive at our main result

$$F = 1 - u + \frac{u(1+v) - \gamma(1+2w)}{2r} + \frac{\gamma}{2r} \cdot \frac{(1+2w)M_{1+w,0}(g) - 2CW_{1+w,0}(g)}{M_{w,0}(g) + CW_{w,0}(g)}. \quad (59)$$

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

Recall that the parameters g, v, w are given by (53), and $u = (1 - y)e^{-\gamma t}$. The constant C in equation (59) is determined from the initial condition, $F(u = 1 - y) = x$, to give

$$C = \frac{-\theta M_{w,0}(\hat{g}) + (1 + 2w)M_{1+w,0}(\hat{g})}{\theta W_{w,0}(\hat{g}) + 2W_{1+w,0}(\hat{g})}. \quad (60)$$

Here we introduced two more shorthand notations

$$\theta = 1 + 2w - \hat{g} + \frac{2r(x - y) + y - 1}{\gamma}, \quad \hat{g} = \frac{(1 - y)v}{\gamma}. \quad (61)$$

The distribution of the total number of cells $\Pi_s(t)$ can be obtained from $G(z, t) = F(z, z, t)$. The survival probability of the cells at time t is

$$S(t) = 1 - F(0, 0, t) \quad (62)$$

where F is given by (59) and (60). Note that in computing $F(x = 0, y = 0, t)$ all the parameters that contain x and y simplify. Setting $x = y = 0$ we get $u = e^{-\gamma t}$, $\theta = 1 + 2w - (v + 1)/\gamma$, and $\hat{g} = v/\gamma$.

From the generating function (59) one can easily extract the clone size distribution $P_{n,m}(t)$ or $\Pi_s(t)$ numerically. Let us start with the simpler total cell distribution $\Pi_s(t)$. The probability $\Pi_s(t)$ is the coefficient of the z^s term in the power series of $G(z, t)$ as given by (37). One way to extract $\Pi_s(t)$ is by using Cauchy's integral formula

$$\Pi_s = \frac{1}{2\pi i} \oint_C \frac{G(z)}{z^{s+1}} dz. \quad (63)$$

Here the contour C goes counterclockwise around the origin in the complex z plane, within the radius of convergence of $G(z)$ (we omitted the time argument for brevity). Consider a contour of a circle of radius R and divide the circle into N equal parts. Now the above integral (63) can be approximated as a sum

$$\Pi_s = \frac{R^{-s}}{N} \sum_{k=0}^{N-1} G\left(\text{Re}^{ik2\pi/N}\right) e^{-iks2\pi/N} \quad (64)$$

which is the discrete Fourier transform scaled by R^{-s}/N . This transformation can be efficiently performed by the fast Fourier transform (FFT) method, which is implemented in most mathematical software. A discussion of this method and approximations of the error terms can be found in [21]. Some care is needed to choose the value of R to avoid numerical problems, as discussed in [21]. Briefly, the contour should be within the radius of convergence of the generating function, but close enough to the closest pole to have a high signal to numerical noise ratio. In our case the choice $R = 1$ was sufficient in all examples we considered. We can check the quality of this numerical method at the special point $\gamma = 1, r = 1/4$, where the explicit solution for $P_{m,n}(t)$ is known (44). For example at $t = 1$, with $N = 32$ the numerical result for $\Pi_s(t)$ differs less than 10^{-15} from the exact expression for $s \leq 31$, and it is precise to at least ten digits for $s \leq 15$.

For the full distribution $P_{m,n}(t)$ one needs two separate contour integrals in both the x and the y planes, which then leads to applying the discrete Fourier transform N^2 times. We have checked the results against the numerical solution of the forward equations (2) and found a perfect agreement (up to about 7 digits). This method has been used to obtain our figures 3 and 4 for $\Pi_s(t)$ and $P_{m,n}(t)$, respectively. In [1] the initial cell is considered to

be A or B with certain probabilities. The corresponding probability distribution is then a simple linear combination of the distribution $P_{m,n}(t)$ we just obtained and the trivial distribution resulting from a single initial B cell (18). Since B cells just die at a fixed rate, their only effect (apart from $P_{0,0}$ and $P_{1,0}$) is to rescale $P_{m,n}(t)$.

Thus we have obtained exact results (59) and (60) for the generating function, which can be easily transformed back to probabilities. Moreover, these exact results simplify in a few special cases (appendix) and in the scaling limit (section 7).

7. Scaling limit

In the large time limit, the distributions $P_{n,m}(t)$ and $\Pi_s(t)$ simplify. Let us consider first the reduced distribution $G(z, t)$. In the large time limit the interesting range of s is $s \sim t$, see e.g. (40), and therefore the interesting range of z is $(1 - z) \sim s^{-1} \sim t^{-1}$. Hence we consider the $t \rightarrow \infty, z \rightarrow 1$ limit with $\zeta = rt(1 - z)/\rho$ kept constant, where $\rho = \gamma/(1 + \gamma)$, see (11).

In order to perform the scaling limit we need the following small argument ($x \ll 1$) limits of the Whittaker functions

$$\begin{aligned} M_{w,0}(x) &= \sqrt{x} - wx^{3/2} + \mathcal{O}(x^{5/2}) \\ W_{w,0}(x) &= -\sqrt{x} \frac{\log x + 2\gamma_E + \psi(1/2 - w)}{\Gamma(1/2 - w)} + \mathcal{O}(x^{3/2}). \end{aligned} \quad (65)$$

Here Γ is the gamma function, $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function, and $\gamma_E = 0.5772\cdots$ is the Euler constant [20]. We also need the identity for the digamma function [19]

$$\psi\left(-\frac{1}{2} - w\right) - \psi\left(\frac{1}{2} - w\right) = \frac{2}{1 + 2w}. \quad (66)$$

Taking the $z \rightarrow 1$ limit of the constant term (60) is particularly easy, since it is independent of time. We find

$$C = C_0(1 - z) + \mathcal{O}(1 - z)^{3/2}, \quad C_0 = r \frac{1 + \gamma}{\gamma^2} \Gamma(1/2 - w). \quad (67)$$

Now we substitute this expression into $F(z, z, t)$ of (59). We also take into account that the variable g , see (53), can be written as

$$g = \frac{uv}{\gamma} = \frac{v\zeta}{r(1 + \gamma)t} e^{-\gamma t}. \quad (68)$$

In the first order of $1/t$ we obtain

$$G(\zeta, t) = 1 - \frac{1}{rt} \cdot \frac{\zeta}{\zeta + 1}. \quad (69)$$

(Some care is needed with terms of type $W_{w,0}(g)$, since e.g. in $\log g = -\gamma t + \log(v\zeta/r(1 + \gamma)t)$ a term proportional to t appears.)

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

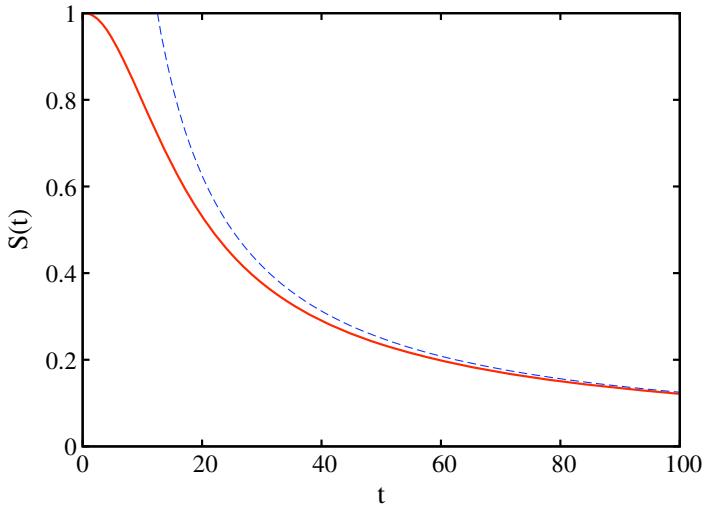


Figure 2. Exact survival probability $S(t) = 1 - F(0, 0, t)$ given by (62), as a function of time at the experimentally measured parameter values $\gamma = 0.28$, $r = 0.08$. The dashed line is the large time asymptotic $1/rt$.

In the scaling limit, $s, t \rightarrow \infty$ with $\mu = \rho s/rt$ kept constant, the generating function $G(\zeta, t)$ of (37) becomes a Laplace transform of Π

$$G(\zeta, t) \rightarrow \frac{rt}{\rho} \int_0^\infty \Pi_s(t) e^{-\zeta\mu} d\mu. \quad (70)$$

Hence we deduce the asymptotic limit of the probability $\Pi_s(t)$ by an inverse Laplace transform [20]

$$\frac{rt}{\rho} \Pi_s(t) \rightarrow \mathcal{L}^{-1}[G(\zeta, t)] = \frac{1}{rt} e^{-\mu} + \left(1 - \frac{1}{rt}\right) \delta(\mu). \quad (71)$$

The first term describes the distribution of the surviving cells, while the second term stands for the extinction of cells. Consequently, the large time survival probability of the population is $1/rt$, or the extinction probability $\Pi_0(t) \sim 1 - 1/rt$. In figure 2 we plotted the exact survival probability $S(t) = 1 - \Pi_0(t)$ of (62) together with the large time asymptotic $1/rt$. The above asymptotic results of course agree with the explicit results in the special point of section 5.

From (71), the regular part of the distribution $\Pi_s(t)$ can be written in a scaling form as

$$\Pi_s(t) = \frac{\rho}{(rt)^2} \exp\left(-\frac{\rho s}{rt}\right) = \frac{\rho}{(rt)^2} \Pi(\mu) \quad (72)$$

with the time independent scaling function

$$\Pi(\mu) = e^{-\mu}. \quad (73)$$

This scaling is demonstrated on figure 3, where the exact expressions (59) for $\Pi_s(t)$ for different times are also depicted. Note that this scaling limit has been already guessed in [1], and derived in [9] in the realm of continuous approximation, that additionally

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

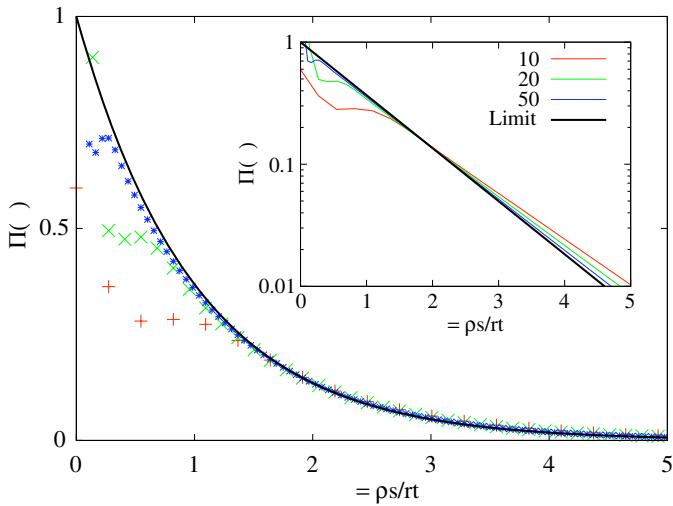


Figure 3. The exact probability $\Pi_s(t)$ of having a total s cells at time t , as given by (59). This probability is depicted at different times in terms of the scaling variable $\mu = \rho s/rt$. The points collapse on the exponential $e^{-\mu}$ limit curve (73). In the inset the same curves are re-plotted on log-scale to emphasize the tail of the distribution. The symbols are the same as in figure 4.

assumed that the B cell population remains ‘slave’ to the A cell population. We will see that the latter approximation is not merely appealing, it is asymptotically correct. This will become evident from the full distribution.

Similarly to the total cell distribution, we can also obtain the scaling limit of the whole $P_{m,n}(t)$ distribution from (59). Taking the $t \rightarrow \infty$ limit while keeping $\xi = rt(1-x)/\rho$ and $\eta = rt(1-y)/\rho$ finite, up to first order in $1/t$ we obtain

$$F(\xi, \eta, t) = 1 - \frac{1}{rt} \cdot \frac{\xi\gamma + \eta}{\xi\gamma + \eta + 1 + \gamma}. \quad (74)$$

Now we need to perform a double inverse Laplace transform to obtain $P_{m,n}(t)$ as a function of m/t and n/t , in the limit $m, n, t \rightarrow \infty$. The extinction probability is again $P_{0,0}(t) = \Pi_0(t) = 1/rt$ in the first order of $1/t$. The probability $P_{m,n}(t)$ for $m, n > 0$ in the scaling limit becomes

$$P_{m,n}(t) = \frac{\gamma}{r^2 t^3} \exp \left[-\frac{\rho(m+n)}{rt} \right] \delta \left(\frac{m - \gamma n}{t} \right). \quad (75)$$

Hence in this limit there are precisely γ times as many A cells as B cells, while the distribution of the cells is given by (72). According to the experiments of Clayton *et al* [1] in skin tissue, where $\gamma = 0.28$, the model predicts about four times more post-mitotic cells than progenitor cells in a clone for large times.

It is possible to give a more detailed description of the cell distribution by taking a different large time limit, namely we need to take the limit $n, m, t \rightarrow \infty$ in such a way

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

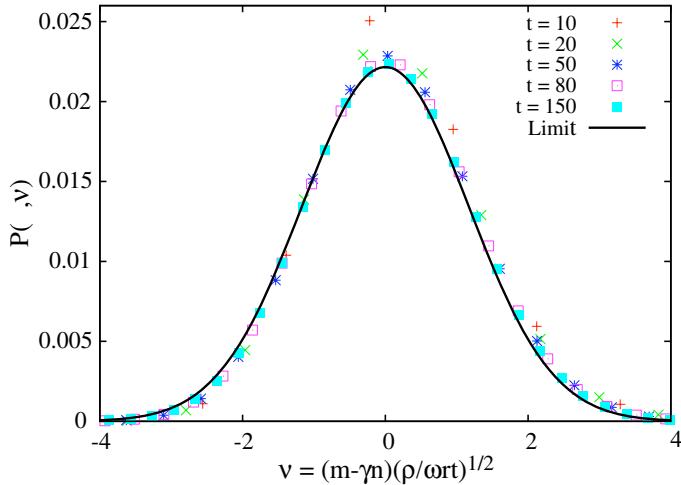


Figure 4. The exact probability $P_{m,n}(t)$ of having m type A , and n type B cells at time t , as given by (59). This probability is depicted at different times in terms of the scaling variable $\nu = (m - \gamma n) \sqrt{\rho/\omega r t}$, at $\mu = (m + n)\rho/rt = \rho/r \approx 2.7$. The points collapse on the Gaussian limit curve (88).

that the following fractions are finite

$$\frac{m}{t} = O(1), \quad \frac{n}{t} = O(1), \quad \frac{m - \gamma n}{t^{1/2}} = O(1). \quad (76)$$

This limit reveals the ‘shape’ of the Dirac delta in (75). This asymptotic limit is of course encoded in the exact results for the generating function (59). Unfortunately, to extract the asymptotic is far from straightforward. Indeed, even from a simple expression for the multivariate generating function, it is usually extremely difficult to extract the asymptotic of the coefficients (let alone the exact expressions for the coefficients). This situation is perhaps surprising, as in the univariate case there are various techniques, the most powerful is the use of complex analysis and the saddle point method. In the multivariate case, the usage of complex methods is much more limited and challenging; for recent progress, see [22] and [23]. In our case, there is an additional difficulty as the explicit expression for the generating function is not a simple rational function, as e.g. in most examples in [23], but it involves the Whittaker functions. Hence instead of extracting the scaling limit from the exact solution, we outline another approach in the next section that also shows an independent way of handling the problem.

8. Fokker–Planck approximation

We shall use a more direct procedure which is however approximate, for instance it does not even provide the asymptotically exact value, $1/rt$, that the clone size is non-zero. However, up to this amplitude one can obtain an expression for the probability distribution $P_{m,n}(t)$ which is typically asymptotically exact in the scaling region (76). The method is essentially the Fokker–Planck or diffusion approximation [17]. One starts with the master equation (2), and treats m, n as continuous variables. This should be valid when $m, n \gg 1$. In this region one can further expand the right-hand side of (2) in the Taylor series to

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

give (we shortly write P instead of $P_{m,n}$)

$$\begin{aligned} (m-1)P_{m-1,n} &= mP - \frac{\partial}{\partial m} mP + \frac{1}{2} \frac{\partial^2}{\partial m^2} mP + \dots \\ mP_{m,n-1} &= mP - \frac{\partial}{\partial n} mP + \frac{1}{2} \frac{\partial^2}{\partial n^2} mP + \dots \\ (m+1)P_{m+1,n-2} &= mP + \frac{\partial}{\partial m} mP - 2 \frac{\partial}{\partial n} mP + \frac{1}{2} \frac{\partial^2}{\partial m^2} mP \\ &\quad - 2 \frac{\partial^2}{\partial m \partial n} mP + 2 \frac{\partial^2}{\partial n^2} mP + \dots \\ (n+1)P_{m,n+1} &= nP + \frac{\partial}{\partial n} nP + \frac{1}{2} \frac{\partial^2}{\partial n^2} nP + \dots. \end{aligned} \tag{77}$$

Using these expansions and ignoring the higher order terms we turn the master equation into a partial differential equation

$$\begin{aligned} \frac{\partial P}{\partial t} &= \gamma P + (\gamma - 2r - m + \gamma n) \frac{\partial P}{\partial n} + 2r \frac{\partial P}{\partial m} \\ &\quad + rm \frac{\partial^2 P}{\partial m^2} - 2rm \frac{\partial^2 P}{\partial m \partial n} + \frac{(1+2r)m + \gamma n}{2} \frac{\partial^2 P}{\partial n^2} \end{aligned} \tag{78}$$

which is the Fokker–Planck equation in our problem.

Let us change m, n to the variables

$$s = m + n, \quad \delta = m - \gamma n. \tag{79}$$

The Fokker–Planck equation becomes

$$\frac{\partial P}{\partial t} = \gamma P + (\gamma - \delta) \frac{\partial P}{\partial s} + [2r + \gamma(\delta + 2r - \gamma)] \frac{\partial P}{\partial \delta} + A \frac{\partial^2 P}{\partial s^2} - 2B \frac{\partial^2 P}{\partial s \partial \delta} + C \frac{\partial^2 P}{\partial \delta^2} \tag{80}$$

with

$$\begin{aligned} A &= \frac{2\gamma s + (1-\gamma)\delta}{2(1+\gamma)}, \quad B = \gamma A, \\ C &= \gamma^2 \frac{(1+2r)(\gamma s + \delta) + \gamma(s-\delta)}{2(1+\gamma)} + r(1+2\gamma) \frac{\gamma s + \delta}{1+\gamma}. \end{aligned} \tag{81}$$

Since $\delta \ll s$ in the scaling region (76), the above coefficients simplify to

$$A = \rho s, \quad B = \gamma \rho s, \quad C = \rho s [r(1+\gamma)^2 + \gamma^2]. \tag{82}$$

We already know the dependence on s , namely $P \sim \exp(-\rho s/rt)$. To determine the dependence on δ we keep only the dominant terms in the Fokker–Planck equation (80) and obtain

$$0 = \gamma P - \gamma \delta \frac{\partial P}{\partial \delta} + C \frac{\partial^2 P}{\partial \delta^2}. \tag{83}$$

Note that all terms in (83) are of the order of P :

$$\delta \frac{\partial P}{\partial \delta} \sim \delta \frac{P}{\delta} \sim P, \quad C \frac{\partial^2 P}{\partial \delta^2} \sim s \frac{P}{\delta^2} \sim P. \tag{84}$$

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

The latter estimate follows from $s \sim \delta^2$ and it actually explains the choice $\delta \sim t^{1/2}$ in the scaling region (76). Note also that the neglected terms from the Fokker–Planck equation (80) are indeed sub-dominant, e.g.

$$A \frac{\partial^2 P}{\partial s^2} \sim s \frac{P}{s^2} = \frac{P}{s}, \quad B \frac{\partial^2 P}{\partial s \partial \delta} \sim s \frac{P}{s \delta} = \frac{P}{\delta}. \quad (85)$$

Solving (83), which is essentially an ordinary differential equation with respect to δ , we find

$$P \sim \exp\left(-\frac{\delta^2}{\omega s}\right), \quad \text{with } \omega = 2(r + \rho^2)(1 + \gamma). \quad (86)$$

Therefore the full scaling solution reads

$$P_{m,n}(t) = \frac{\gamma}{(rt)^2 \sqrt{\pi \omega s}} \exp\left(-\frac{\rho s}{rt} - \frac{\delta^2}{\omega s}\right). \quad (87)$$

The amplitude, including the $1/\sqrt{s}$ factor, is obtained by requiring $\Pi_s = \int P(s, \delta) d\delta / (1 + \gamma)$, using (72). Note that the distribution (87) is normalized as $\int P dm dn = \int P ds d\delta / (1 + \gamma) = (rt)^{-1}$.

The limit distribution (87) can be written in a scaling form

$$P_{m,n}(t) = \frac{\gamma}{(rt)^{5/2}} \sqrt{\frac{\rho}{\omega}} P(\mu, \nu), \quad \text{with } P(\mu, \nu) = \frac{e^{-\mu-\nu^2/\mu}}{\sqrt{\pi \mu}} \quad (88)$$

with scaling variables

$$\mu = \frac{\rho s}{rt}, \quad \nu = \delta \sqrt{\frac{\rho}{\omega rt}}. \quad (89)$$

This scaling is probed in figure 4, using exact values for $P_{m,n}(t)$ from (59). One can see that the scaling limit (87) provides an excellent approximation already for times $t \gtrsim 10$, and the finite time curves converge to the scaling function (88). Note also that in the special point $\gamma = 1, r = 1/4$ the distribution (44) converges exactly to the scaling limit (88).

9. Discussion

We investigated a specific stochastic process (1) that has been proposed [1] to describe cell maintenance measurements of murine tail epidermis. The chief ingredient of the stochastic process (1) is the self-duplication and differentiation of the progenitor cells without measurable contribution from stem cells. (Stem cells activate during repair from severe injuries.) The same mechanism apparently underlies the maintenance of pancreatic islets [7] and lung homeostasis [8].

We derived an exact solution for a two-type branching process in terms of generating functions. This generating function provides the exact probability distribution valid for all times. Several features, e.g. the probability the clone has disappeared, admit compact exact expressions in terms of the special functions, while the probability distribution can be computed very accurately by an easy numerical inverse transformation.

The stochastic process (1) is very simple and some of the underlying assumptions deserve further scrutiny. For instance, the time between cell divisions is exponentially

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

distributed, which is considered unrealistic in many situations [24, 25]. The authors of [9] showed, however, that an exponentially distributed cell division time fits well the experimental results [1] for all times apart from the very early regime; the same model (1) with other cell division time distributions appears to be less accurate. Another prominent feature of our analysis is the disregard of spatial characteristics. In the context of epidermis, one might want to consider the two-dimensional version of the two-type branching process (1). It is, however, quite challenging to introduce the right spatial extension when modeling such a soft tissue as the skin. One type of spatial model is partly amenable to analysis [10] due to an intimate connection with models of voting and monomer–monomer catalytic reactions [26]–[28]. Intriguingly, the model presented in this paper, which completely disregards real space, already provides an excellent fit to experimental data. Despite such a success, spatial characteristics should eventually matter, and their role deserves further investigation.

An exact solution of the specific two-type branching process (1) raises the hope that other two-type branching processes could be amenable to exact treatments. Some two-type branching processes have been suggested in the context of tumor formation, see e.g. [29]–[31]. Indeed, cancer often arises when a progenitor cell undergoes a series of mutations in a way that the proliferation of a mutant clone dominates the differentiation or death [32]–[36]. The two-type branching model of cancer is indeed tractable [37]. The complication is that cancer typically involves multiple mutations [38]–[43], so the quantitative description may require a multiple-type branching process.

Acknowledgments

We are grateful for financial support from NSF grant CCF-0829541(PLK), the John Templeton Foundation, the NSF/NIH grant R01GM078986, and J Epstein (TA).

Appendix: Special cases

As a check of self-consistency it is useful to extract the explicit results of section 5 from the general approach of section 6. At the special point $\gamma = 1, r = 1/4$ the potential in equation (52) is purely quadratic $V = 1/4u^2$, and the solution of (52) becomes $Z = \sqrt{u}(C + \log u)$. After transforming $Z(u)$ back to $F(t)$ and fitting to the boundary conditions, we recover (36).

Exact results (59) and (60) also simplify on a few lines in the parameter space.

A.1. Horizontal line $r = 1/4$

On this line the u -independent term in the potential in equation (52) vanishes and the Schrödinger equation becomes

$$Z'' + \frac{1}{4} \left(\frac{\gamma - 1}{\gamma^2} u^{-1} + u^{-2} \right) Z = 0. \quad (\text{A.1})$$

In the $u \rightarrow 0$ limit, $Z(u)$ behaves as \sqrt{u} . This suggests choosing $k = \sqrt{u}$ as the basic variable and seeking a solution proportional to k . Hence we write

$$Z(u) = kG(k), \quad u = \frac{\gamma^2}{1 - \gamma} k^2. \quad (\text{A.2})$$

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

The amplitude $\gamma^2/(1-\gamma)$ has been chosen to get rid off γ in the coefficients of the governing equation for $G(k)$:

$$\frac{d^2G}{dk^2} + \frac{1}{k} \frac{dG}{dk} - G = 0. \quad (\text{A.3})$$

Solutions to this equation are linear combinations of the modified Bessel function $I_0(k)$ and $K_0(k)$, i.e.

$$G(k) = C_1 I_0(k) + C_2 K_0(k). \quad (\text{A.4})$$

Then the function $F(u)$ is given by

$$F = 1 + u + 2\gamma k \frac{CI_1(k) - K_1(k)}{CI_0(k) + K_0(k)} \quad (\text{A.5})$$

where we used identities $I'_0(x) = I_1(x)$, $K'_0(x) = -K_1(x)$ and $C = C_1/C_2$. We can re-write this as

$$F(x, y, t) = 1 + (1-y)e^{-\gamma t} + 2\gamma k \frac{CI_1(k) - K_1(k)}{CI_0(k) + K_0(k)} \quad (\text{A.6})$$

with

$$k = \gamma^{-1} e^{-\gamma t/2} \sqrt{(1-\gamma)(1-y)} \quad (\text{A.7})$$

where $C = C(x, y)$ is determined by matching to the initial condition $F(x, y, t=0) = x$. One gets

$$C = \frac{\gamma\kappa K_1(\kappa) - (1-(x+y/2)) K_0(\kappa)}{\gamma\kappa I_1(\kappa) + (1-(x+y/2)) I_0(\kappa)} \quad (\text{A.8})$$

with

$$\kappa = k(t=0) = \gamma^{-1} \sqrt{(1-\gamma)(1-y)}. \quad (\text{A.9})$$

A.2. Special curve $\gamma = 2r/(1-2r)$

On this curve $w = 0$, hence the term proportional to g^{-1} vanishes in the Schrödinger equation (54), which then can be solved in terms of Bessel functions. Here instead, we derive the simplified form from the general solution (54). For this we need some limit properties [20] of the Whittaker functions

$$\begin{aligned} M_{0,0}(g) &= \sqrt{g} I_0(g/2) \\ W_{0,0}(g) &= \sqrt{g/\pi} K_0(g/2) \\ M_{1,0}(g) &= (1-g)\sqrt{g} I_0(g/2) + g^{3/2} I_1(g/2) \\ W_{1,0}(g) &= \frac{1}{2\sqrt{\pi}} [-(1-g)\sqrt{g} K_0(g/2) + g^{3/2} K_1(g/2)]. \end{aligned} \quad (\text{A.10})$$

By using these expressions in (59), we obtain

$$F(u) = 1 - u + \frac{u}{2r} \left[1 + v \cdot \frac{I_1(g/2) - CK_1(g/2)}{I_0(g/2) + CK_0(g/2)} \right] \quad (\text{A.11})$$

with constant

$$C = \frac{-\chi I_0(\hat{g}/2) + v I_1(\hat{g}/2)}{\chi K_0(\hat{g}/2) + v K_1(\hat{g}/2)}, \quad \chi = \frac{2r(x-y)}{1-y} - 1. \quad (\text{A.12})$$

References

- [1] Clayton E, Doupé D P, Klein A M, Winton D J, Simons B D and Jones P H, *Single type of progenitor cell maintains normal epidermis*, 2007 *Nature* **446** 185
- [2] Clarke J D and Tickle C, *Fate maps old and new*, 1999 *Nat. Cell Biol.* **1** E103
- [3] Jonkers J and Berns A, *Conditional mouse models of sporadic cancer*, 2002 *Nat. Rev. Cancer* **2** 251
- [4] Blanpain C and Fuchs E, *Epidermal homeostasis: a balancing act of stem cells in the skin*, 2009 *Nat. Rev. Mol. Cell Biol.* **10** 207
- [5] Potten C S, *Epidermal proliferative unit: the possible role of central basal-cell*, 1974 *Cell Tissue Kinet.* **7** 77
- [6] Jones P and Simons B D, *Epidermal homeostasis: do committed progenitors work while stem cells sleep?*, 2008 *Nat. Rev. Mol. Cell Biol.* **9** 83
- [7] Dor Y, Brown J, Martinez O I and Melton D A, *Adult pancreatic β -cells are formed by self-duplication rather than stem-cell differentiation*, 2004 *Nature* **429** 41
- [8] Giangreco A et al, *Stem cells are dispensable for lung homeostasis but restore airways after injury*, 2007 *Proc. Natl. Acad. Sci.* **106** 9286
- [9] Klein A M, Doupé D P, Jones P H and Simons B D, *Kinetics of cell division in epidermal maintenance*, 2007 *Phys. Rev. E* **76** 021910
- [10] Klein A M, Doupé D P, Jones P H and Simons B D, *Mechanism of murine epidermal maintenance: cell division and the voter model*, 2008 *Phys. Rev. E* **77** 021910
- [11] Pharr P N et al, *A stochastic model for mast cell proliferation in Culture*, 1985 *J. Cell. Physiol.* **125** 379
- [12] Vogel H, Niewisch H and Matioli G, *Stochastic development of stem cells*, 1969 *J. Theor. Biol.* **22** 249
- [13] Kimmel M and Axelrod D E, 2002 *Branching Processes in Biology* (New York: Springer)
- [14] Dingli D, Traulsen A and Pacheco J M, *Compartmental architecture and dynamics of hematopoiesis*, 2007 *PLoS ONE* **2** e345
- [15] Dingli D, Antal T, Traulsen A and Pacheco J M, *Progenitor cell self-renewal and cyclic neutropenia*, 2009 *Cell Prolif.* **42** 330
- [16] Athreya K B and Ney P E, 2004 *Branching Processes* (Mineola, New York: Dover)
- [17] van Kampen N G, 1997 *Stochastic Processes in Physics and Chemistry* 2nd edn (Amsterdam: North-Holland)
- [18] Logan J D, 2008 *An Introduction to Nonlinear Partial Differential Equations* (New Jersey: Wiley)
- [19] Bender C M and Orszag S A, 1978 *Advanced Mathematical Methods for Scientists and Engineers* (New York: McGraw-Hill)
- [20] Gradshteyn I S and Ryzhik I M, 2007 *Table of Integrals, Series, and Products* 7th edn (Amsterdam: Elsevier)
- [21] Cavers J K, *On the fast Fourier inversion of probability generating functions*, 1978 *J. Inst. Maths Appl.* **22** 275
- [22] Flajolet P and Sedgewick R, 2009 *Analytic Combinatorics* (New York: Cambridge University Press)
- [23] Pemantle R and Wilson M C, *Twenty combinatorial examples of asymptotics derived from multivariate generating functions*, 2008 *SIAM Rev.* **50** 199
- [24] Bailey N T J, 1964 *The Elements of Stochastic Processes* (New York: Wiley)
- [25] Oprea M and Kepler T B, *Improved inference of mutation rates; II. Generalization of the Luria–Delbrück distribution for realistic cell-cycle time distributions*, 2001 *Theor. Popul. Biol.* **59** 49
- [26] Krapivsky P L, *Kinetics of monomer–monomer surface catalytic reactions*, 1992 *Phys. Rev. A* **45** 1067
- [27] Frachebourg L and Krapivsky P L, *Exact results for kinetics of catalytic reactions*, 1996 *Phys. Rev. E* **53** R3009
- [28] Liggett T M, 1999 *Stochastic Interacting Systems: Contact, Voter, and Exclusion Processes* (New York: Springer)
- [29] Iwasa Y, Michor F, Komarova N L and Nowak M A, *Population genetics of tumor suppressor genes*, 2005 *J. Theor. Biol.* **233** 15
- [30] Iwasa Y, Nowak M A and Michor F, *Evolution of resistance during clonal expansion*, 2006 *Genetics* **172** 2557
- [31] Kendall D G, *Birth-and-death processes, and the theory of carcinogenesis*, 1960 *Biometrika* **47** 13
- [32] Attolini S-O C and Michor F, *Evolutionary theory of cancer*, 2009 *Ann. NY Acad. Sci.* **1168** 23
- [33] Dingli D, Traulsen A and Michor F, *(A)Symmetric stem cell replication and cancer*, 2007 *PLOS Comput. Biol.* **3** e53
- [34] Fearon E R, Hamilton S R and Vogelstein B, *Clonal analysis of human colorectal tumors*, 1987 *Science* **238** 193
- [35] Fearon E R and Vogelstein B, *A genetic model for colorectal tumorigenesis*, 1990 *Cell* **61** 759

Exact solution of a two-type branching process: clone size distribution in cell division kinetics

- [36] Nowak M A, 2006 *Evolutionary Dynamics: Exploring the Equations of Life* (Cambridge, MA: Belknap Press)
- [37] Antal T and Krapivsky P L, *Models of tumor progression*, 2010 in preparation
- [38] Armitage P and Doll R, *The age distribution of cancer and a multi-stage theory of carcinogenesis*, 1954 *Brit. J. Cancer* **8** 1
- [39] Beerenwinkel N et al, *Genetic progression and the waiting time to cancer*, 2007 *PLOS Comput. Biol.* **3** e225
- [40] Durrett R and Mayberry J, *Traveling waves of selective sweeps*, 2009 arXiv:0910.5730
- [41] Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, Karchin R, Vogelstein B and Nowak M A, *Accumulation of driver and passenger mutations during tumor progression*, 2009 arXiv:0912.1627
- [42] Durrett R and Moseley S, *Evolution of resistance and progression to disease during clonal expansion of cancer*, 2010 *Theor. Popul. Biol.* **77** 42
- [43] Luebeck E G and Moolgavkar S H, *Multistage carcinogenesis and the incidence of colorectal cancer*, 2002 *Proc. Nat. Acad. Sci.* **99** 15095