

# Branching Stochastic Processes with Immigration in Analysis of Renewing Cell Populations

Andrei Yakovlev<sup>a,\*</sup> and Nikolai Yanev<sup>a,b</sup>

<sup>a</sup>*Department of Biostatistics and Computational Biology, University of Rochester, 601 Elmwood Avenue, Box 630, Rochester, New York 14642, U.S.A.*

<sup>b</sup>*Department of Probability and Statistics, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria*

## Abstract

This paper considers the utility of a new class of stochastic branching processes with non-homogeneous immigration in modeling complex renewing cell systems. Such systems typically include the population of stem cells that provides an inexhaustible supply of cells necessary for maintaining the cellular composition of a tissue. A stem cell may be induced to transform (differentiate) into a progenitor cell. Progenitor cells retain the ability to proliferate and their function is believed to provide a quick proliferative response to an increased demand for cells in the population. There may be several sub-types of progenitor cells. Terminally differentiated cells do not divide under normal conditions; they are responsible for maintaining tissue-specific functions. Recent advancements in experimental techniques offer considerable scope for quantitative studies of *in vivo* cell kinetics based on stochastic modeling of renewing cell populations. However, no ready-made theory is currently available to take full advantage of these advancements. This paper introduces such a theory with a special focus on its feasibility in biological applications.

*Key Words:* branching processes, immigration, cell kinetics, stem cells, progenitor cells

*\* Corresponding author:* Andrei Yakovlev, Department of Biostatistics and Computational Biology, University of Rochester, 601 Elmwood Avenue, Box 630, Rochester, New York 14642, ph: (585)275-6688, fax: (585)273-1031  
e-mail: Andrei\_Yakovlev@urmc.rochester.edu

# 1. Introduction

The present paper deals with a design of biological experiments that have never been conducted. This situation is quite common in physics but much less so in mathematical biology. This is the reason why we are currently unable to illustrate the proposed methodology with a specific application to experimental data analysis. The need for such experiments will become clear once biological investigators have realized the power of mathematical methods in studies of complex multi-type cell systems. This is especially true for studies of cell proliferation and differentiation *in vivo* where methods of mathematical modeling and statistical analysis do not keep abreast with the development of experimental techniques. What is now required is to demonstrate that the most basic problems of biological interest in cell kinetics are mathematically approachable.

Recent decades have seen impressive advancements in experimental approaches to *in vivo* cell kinetics. These advancements have made it possible to distinguish many cell types by antibody labeling so that cells of different types can be counted in the dissociated tissue by using flow cytometry. In other words, it is now possible to make snapshots of multi-type cell populations as they develop in time. The ability to distinguish and count millions of cells opens up a new research area to the biomathematician. The main focus in this area should be on extracting the necessary information on unobservable but biologically meaningful parameters of population dynamics from observable cell counts. It is easy to envision numerous experimental settings where model-based quantitative inference is very much in need.

To illustrate the latter statement, consider a hypothetical renewing cell population consisting of three types of cells and depicted in Figure 1. The first compartment is represented by stem cells. Such cells are poorly defined and found at relatively low frequency, thus making their experimental analysis difficult. It is fair to say that stem cells are not accessible to direct observation as far as their kinetics and cell cycle structure are concerned. Even their total counts at a particular anatomical site are not available for quantitative analysis. The immediate downstream progeny of stem cells, the progenitor cells, can typically be identified with cell surface markers. The same is true for terminally differentiated cells comprising the third compartment of the three-type cell population presented in Figure 1. The main function of stem cells is to provide an inexhaustible supply of cells in order to maintain the cellular composition of a tissue. A stem cell (SC) remains in a “resting” state until being induced to divide or transform (differentiate) into a progenitor cell (PC). The latter cell type is partially committed to further differentiation into another cell type but it retains the ability to proliferate. The main function of progenitor cells is believed to provide a quick proliferative response to an increased demand for cells in the population. Terminally differentiated (mature) cells typically represent a final cell type; they are responsible for maintaining tissue-specific

functions and they do not divide under normal conditions. All cell types are susceptible, in variable degrees, to death. Two specific biological examples of such cell systems are given in the next section.

Leaving aside statistical aspects of the problem, the most important point here is that modern experimental techniques yield counts of PCs and mature cells as functions of time. These counts per se provide little information on the underlying biological processes. The transition rates, denoted by  $r_1, r_2, r_3, r_4$  in Figure 1, represent far more informative characteristics. Especially interesting is the rate  $r_1$  that contains information about the functioning of the population of stem cells. Reconstructing the influx of PCs furnishes the possibility to study the behavior of unobservable stem cells by the indirect route. Also, it would be desirable to learn more about PCs in terms of probabilistic characteristics of the duration and possible outcomes of their mitotic cycle. All this potentially useful information is inaccessible without mathematical modeling. Up to date, there are no general modeling techniques that are adequate for the complexity of cell kinetics in renewing tissues. The present paper considers the utility of branching stochastic processes for building realistic models of *in vivo* cell kinetics, provided the necessary experimental data are available. While the proposed models are much more sophisticated than those traditionally used by practitioners, they meet the requirements of feasibility and usefulness.

The theory of branching stochastic processes has proven itself a powerful tool for cell kinetics studies. This theory has its origin in the pioneering works of Bienaymé [1] and Galton and Watson [2] motivated by applications in genealogy. The term “branching processes” was introduced by A. N. Kolmogorov in 1940-ies. It is noteworthy that the first asymptotic result in the theory of branching processes was obtained by Kolmogorov in the context of a specific biological problem [3]. Age-dependent branching processes were first considered by Bellman and Harris [4, 5], also in relation to biological problems. The first branching process with immigration (Markov process with homogeneous Poisson immigration) was studied by Sevastyanov [6]. A comprehensive account of fundamental achievements in the theory of branching processes is given in many excellent books by Harris [7], Sevastyanov [8], Mode [9], Athreya and Ney [10], Jagers [11], Assmussen and Hering [12], Guttorp [13], and Kimmel and Axelrod [14]. A special focus on biological applications is characteristic of the books by Jagers [11], Yakovlev and Yanev [15], and Kimmel and Axelrod [14].

In the present paper, we introduce and investigate new models of multi-type age-dependent branching processes with non-homogeneous in time immigration as pertinent models of renewing cell populations. These processes represent building blocks for designing stochastic models of cell systems of increasing complexity. They furnish a universal modeling technique for the processes of cell proliferation and differentiation in various experimental settings. To the best of our knowledge, the branching stochastic processes with non-homogeneous Poisson or re-

new type immigration have never been studied in the theory of branching processes. Such models are mathematically challenging and yet tractable. The relevant results are given in Appendix 1. In addition, new characteristics of branching processes, such as the residual life-time distribution and its limiting form, are introduced and investigated (see Appendix 2) in relevance to labeling experiments. The original theoretical results presented in this paper provide the basis on which parametric inferential procedures could be built.

The paper is organized as follows. We present two relevant biological examples in Section 2. Section 3 deals with models of the influx of PCs and their subsequent evolution within the framework of branching stochastic processes with immigration. The main purpose of Section 3 is to demonstrate that the influx of PCs can be reconstructed from moments of the number of PCs as functions of time. The models discussed in this section are general enough to provide a realistic description of *in vivo* cell kinetics. However, it is unclear whether they are identifiable when only cell counts are available for analysis. To surmount this difficulty, we consider a similar modeling technique for labeling experiments which provide an additional source of information on parameters of the mitotic cycle. This technique is presented in Section 4. Section 5 discusses the potentialities of the proposed methodology in biological applications.

## 2. Examples of Renewing Cell Systems

### 2.1. Development of Oligodendrocytes

The precursor cell that gives rise directly to oligodendrocytes was first discovered by Raff, Miller, and Noble in 1983 [16], when it was named as an oligodendrocyte/type-2 astrocyte (O-2A) progenitor cell for the two cell types it could generate *in vitro*. As it remains uncertain whether type-2 astrocytes are generated during normal development of the central nervous system (CNS) *in vivo* [17], this cell is also referred to as an oligodendrocyte precursor cell (OPC), and will be referred to herein as an O-2A/OPC. Such cells appear to be present in all regions of the perinatal rat CNS (except possibly the olfactory bulb [18]), and cells with similar properties also have been isolated from the human CNS [19]. O-2A/OPCs have proven a remarkably useful population for studies on development of the CNS due to the ability of these cells to grow readily in clonal culture and due to the ease with which oligodendrocytes and progenitor cells can be distinguished from each other by visual inspection.

As an experimental system, the rat spinal cord is well suited for studying its normal and pathological development *in vivo* by methods of mathematical modeling and simulation. The whole spinal cord can be removed from embryos and dissociated into single cells so that the total cell count is available. In the developing spinal cord, the O-2A/OPC appears to be itself

generated from an earlier progenitor cell that also is restricted to the generation of glia. As shown in Figure 2, immediately prior to the O-2A/OPC in development is a tripotential glial-restricted precursor (GRP) cell that can be directly isolated from the developing spinal cord [20]. This cell type appears in the spinal cord several days earlier than the first appearance of O-2A/OPCs in this tissue [21]. The GRP cell type was discovered by Rao and Mayer-Proschel in 1997 [22].

The above-described cell types can be identified by their antigenic phenotype. Oligodendrocytes express galactocerebroside (GalC), a myelin-specific glycolipid. GalC+ oligodendrocytes are first found in the rat spinal cord at E20 (day 20th of embryonic life). O-2A/OPCs are labeled with the O4 monoclonal antibody but do not yet express GalC (O4+/GalC- phenotype). There is also an earlier stage of O-2A/OPC development at which these cells are labeled with the A2B5 monoclonal antibody, but are not yet O4+. GRP cells are labeled with the A2B5 antibody. Once O-2A/OPCs are generated, this labeling alone no longer allows unambiguous identification of GRP cells. However, those cells that are A2B5+/O4+ can be defined as O-2A/OPCs. Clonal analysis indicates that the A2B5+ cells isolated from the rat spinal cord at E16 or earlier are all GRP cells. GRP cells first appear at E12.

The GRP cell is itself derived from a totipotent neuroepithelial stem cell (NSC), which can give rise to all of the cell types of the CNS [22]. The NSC gives rise to glia by first giving rise to GRP cells. Analogously, the NSC gives rise to neurons by first generating a neuron-restricted precursor (NRP) cell that gives rise to multiple classes of neurons but not to glial cells [23]. There is ample experimental evidence supporting the hypothesis of asymmetric division of multipotent stem cells that generate diverse cell types in the CNS [24, 25, 26, 27, 28, 29, 30]. Experimental studies on both vertebrate animals and invertebrates indicate a crucial role of the membrane-associated protein NUMB in this process [30]. The outcome of the mitotic cycle of asymmetrically dividing cells is exactly one cell of the same type with another daughter cell undergoing differentiation into a different cell type. The asymmetric division is a mechanism by which stem cells prevent their extinction. This is the reason why the asymmetric division of stem cells is a key element of the model presented in Figure 2. In contrast, the GRP cells, as well as O-2A/OPCs, are likely to divide symmetrically, i.e., their mitotic cycle results in two cells of the same type. However, the presence of some asymmetric divisions in these populations of cells cannot be ruled out.

The ability to study both normal and pathological development of the CNS *in vivo* is hindered by the lack of analytical tools that are needed to extract the desired information on unobservable parameters of cell kinetics. A stochastic model of the *in vitro* development of oligodendrocytes has been developed and tested against experimental data in several publications [31, 32, 33, 34, 35, 36, 37, 38, 39]. These papers employ a multi-type Bellman-Harris branching process to model proliferation of O-2A/OPCs and their transformation into termi-

nally differentiated oligodendrocytes. The model was based on the following assumptions:

A1. The process starts with a single progenitor cell of type 1 at time 0. The time to division or differentiation of this initiator cell is assumed to be a non-negative random variable (r.v.)  $T_1$  with cumulative distribution function (c.d.f.)  $F(x) = P(T_1 \leq x)$  satisfying the condition:  $F(0) = 0$ .

A2. After completion of its mitotic cycle, every progenitor cell of type  $l \geq 1$  either divides producing two new progenitor cells of age 0 and type  $l + 1$  with probability  $p_l$ , or transforms into a differentiated cell of type  $l = 0$  (oligodendrocyte) with probability  $1 - p_l$ . The probabilities  $p_l$  can be described by an arbitrary function of the mitotic cycle label  $l$  that satisfies the natural constraints:  $0 \leq p_l \leq 1$  for all  $l \geq 1$ . In [32], these probabilities are specified as  $p_l = p + qr^l$ , where  $p$ ,  $q$  and  $r$  are free parameters with  $p$  representing the limiting probability of division of progenitor cells as the number of cycles tends to infinity.

A3. The times to division or differentiation of progenitor cells of types  $l \geq 2$  are identically distributed non-negative r.v.'s  $T_l$  with c.d.f.  $G(x) = P(T_l \leq x)$  satisfying the condition:  $G(0) = 0$ . This distribution is generally distinct from  $F(x)$ .

A4. Among the two cell types, oligodendrocytes appear to be more susceptible to death. In the culture conditions used in these studies, the death of oligodendrocytes normally begins on day 7 after plating and the rate of death increases with time. Furthermore, as our independent observations show, the proportion of cells (both progenitor cells and oligodendrocytes) surviving by day 8 is no more than 80%. Differentiated oligodendrocytes neither divide nor differentiate further, but they may die; their lifespan  $T_0$  has c.d.f.  $L(x) = P(T_0 \leq x)$ .

A5. Whenever the counts of dead oligodendrocytes are utilized for estimation purposes, the model needs to be extended further to include the following assumption: every dead oligodendrocyte disappears (disintegrates) from the field of observation after a random lapse of time  $T_{-1}$  distributed in accordance with c.d.f.  $H(x) = P(T_{-1} \leq x)$ . The time to the event of disintegration is expected to be quite long, as there are no macrophages present in the culture to clear away cell debris.

A6. The cells do not migrate out of the field of observation (confirmed by time-lapse video records).

A7. The usual independence assumptions are adopted.

Assumption A3 was relaxed in [35] by allowing for non-identical distribution of the mitotic cycle duration and the time to differentiation. The results of data analysis corroborate the idea that the two r.v.'s are not identically distributed, which is why we have accommodated this plausible assumption in the proposed general methodology as discussed in the subsequent sections. One of the important byproducts of the previous work was that the popular clock model of oligodendrocyte generation *in vitro* [40] was disproved by testing a more general (hierarchical) model against experimental data [36, 38]. Several applications of this stochastic

model in combination with new estimation techniques have provided a better quantitative insight into the processes underlying the generation of oligodendrocytes *in vitro*. However, the methodology is not directly applicable under *in vivo* conditions because the processes of embryonic cell development *in vivo* involve immigration of stem cells to the GRP cell compartment. Unlike the previous *in vitro* model, where the number of initiator cells at time  $t = 0$  was fixed, we now need to consider the influx of cells into the GRP cell population. This influx is unobservable, but we have to reconstruct it in order to describe the dynamics of progenitor cells (GRPs and O-2A/OPCs) and oligodendrocytes. Reconstructing the transition between stem cells and GRP cells is also needed to determine whether or not the population of stem cells is affected in the course of such developmental maladies as those caused by iron deprivation and hypothyroidism. The development of cells within each compartment is governed by the probabilities of cell division, differentiation, and death, as well as the mitotic cycle parameters. Specific numerical values of these parameters remain unknown.

## 2.2. Kinetics of Leukemic Cells

Malignant stem cells have now been described for cancers of the blood, breast and brain. In each case, the unifying feature of such cancers is a unique subpopulation of stem cells that are responsible for tumor genesis and perpetuation [41, 42, 43]. Historically, drug development for cancer has not considered the existence of tumor heterogeneity, particular with regard to this minor subset of essential tumor-initiating stem cells. While it is well recognized that stem cells lie at the origin of tumor growth and represent a critical target for therapeutic intervention, the effects of existing treatment regimens on the population of tumor stem cells are largely unknown. Thus, understanding how any given drug regimen modulates the growth, differentiation, and survival of cancer stem cells is arguably the most important research avenue in cancer chemotherapy.

The initiation and perpetuation of leukemia derives from the leukemic stem cell (LSC) population. The propagation of leukemic disease is critically dependent on the immediate downstream progeny of LSC, the leukemic progenitor (LP) population. Cells of this nature can easily be screened using common *in vitro* colony assays and can also be identified with cell surface markers using flow cytometry. The LP cells give rise to blast cells (LB) representing a final cell type in the development of primitive leukemic cells. Blast cells do not differentiate but they may die. The evolution of the initial LSC is shown in Figure 3. Since the asymmetric mitosis seems to be a general mechanism by which stem cells maintain their population (see Section 2.1), it is highly plausible that LSC divide asymmetrically. We proceeded from this assumption when drawing the scheme shown in Figure 3.

Virtually nothing is known about quantitative characteristics of the transition process from

LSC to LP and then to LB. Each transition step is determined by such unobservable characteristics as the probabilities of cell division, differentiation, and death, as well as the mitotic cycle parameters. All the important parameters of cell kinetics in leukemia, including the rate of transition of stem cells to the LP compartment, have not been characterized. The situation here is very similar to that in the example discussed in Section 2.1. The problem becomes even more challenging when studying normal hematopoiesis because the corresponding model structure involves many more distinct cell types.

### 3. A General Model

From a modeling standpoint, the most serious complication brought to the fore by the examples of Section 2 is that the *in vivo* cell kinetics involve immigration of stem cells to the compartment of progenitor cells, henceforth denoted by PCs. Therefore, we will explore a wide class of multi-type age-dependent branching stochastic processes with immigration as a pertinent model of the processes under consideration. Although some notable results are presented in [44], no general theory is currently available for the multi-type decomposable branching processes so that every particular case is of theoretical interest.

In what follows, the newly formed (as a result of stem cell differentiation) PCs will be termed “immigrants” or “arrivals”. These arrivals do not include those resulting from either symmetric or asymmetric mitotic divisions of the progenitor cells. Then we can define a stochastic process  $\mu(t) = \{\# \text{ arrivals by time } t\}$ . Upon arrival, every PC enters the mitotic cycle (MC) with probability  $p$  or undergoes differentiation (culminating in the transition to the next population of PCs of a different type or of mature cells) with probability  $1 - p$ . This mode of transition to the differentiation pathway is suggested by our studies of the kinetics of O-2A/OPCs *in vitro* [35]. The outcomes of the MC include mitotic division into two daughter cells of the same type as well as cell death or exit from the state of active proliferation. The division probability is denoted by  $\beta$ . The MC duration has c.d.f.  $G$ , while the time to the event of differentiation (transition) is distributed in accordance with c.d.f.  $L$ .

#### 3.1. An Age-dependent Branching Process of Progenitor Cell Evolution without Immigration

We begin by considering a fairly general model of a population of progenitor cells (PCs) without immigration. This is the necessary step on the road to its extension allowing for immigration processes. First, we have to describe the evolution of an individual PC from its birth to the moment it leaves the population under consideration.

### 3.1.1. The Evolution of PCs

Every PC with probability  $p$  has a random lifetime  $\xi$  with c.d.f.  $G(x) = \Pr(\xi \leq x)$  or with probability  $1 - p$  it differentiates into another cell type. In other words, the lifetime  $\xi$  is identical to the MC duration. At the end of its life, every PC gives rise to  $\nu$  offsprings (of the same cell type) with discrete distribution  $p_k = \Pr(\nu = k)$ ,  $\sum_{k=0}^{\infty} p_k = 1$ . It takes a random time  $\eta$  with c.d.f.  $L(x) = \Pr(\eta \leq x)$  for the event of differentiation to actually occur. The random variable (r.v.)  $\nu$  will be characterized by the so-called offspring probability generating function (p.g.f.)  $h(s) = Es^\nu = \sum_{k=0}^{\infty} p_k s^k$ ,  $|s| \leq 1$ . The most representative example is given by  $h(s) = 1 - \beta + \beta s^2$ ,  $|s| \leq 1$ , implying that a PC divides with probability  $\beta$  or dies with probability  $1 - \beta$ . In what follows, however, we will consider the p.g.f.  $h(s)$  of the general form.

### 3.1.2. The Model

Let the stochastic processes  $Z(t)$  and  $Z(t, x)$  represent the total number of PCs and the number of PCs of age  $\leq x$  at time  $t \geq 0$ , respectively. Let  $F(t; s) = E\{s^{Z(t)} | Z(0, 0) = 1\}$  be the p.g.f. of the number of cells  $Z(t)$  at time  $t$  produced by one cell of zero age and let  $F(t, x; s) = E\{s^{Z(t, x)} | Z(0, 0) = 1\}$  be the corresponding p.g.f. for the process  $Z(t, x)$ . Note that if  $x \geq t$  then  $Z(t) = Z(t, x)$ . It is also worth introducing a stochastic process  $\bar{Z}_t(y)$  to represent the number of PCs at time  $t$  with the residual lifetime greater than  $y$ . The latter process will be recalled in Section 3.5 as it plays a key role in the analysis of labeling experiments.

By conditioning on the evolution of the first (initiator) cell and applying the law of the total probability (LTP), one can derive the following integral equations for  $F(t; s)$  and  $F(t, x; s)$  (see, e.g., [7] for relevant techniques):

$$F(t; s) = p \int_0^t h(F(t-u; s)) dG(u) + s\{p[1 - G(t)] + (1-p)[1 - L(t)]\} + (1-p)L(t), \quad (1)$$

$$F(t, x; s) = p \int_0^t h(F(t-u, x; s)) dG(u) + [s\delta(x-t) + 1 - \delta(x-t)]\{p[1 - G(t)] + (1-p)[1 - L(t)]\} + (1-p)L(t), \quad (2)$$

with the initial conditions:  $F(0; s) = F(0, 0; s) = s$ , where  $\delta(z) = 1$  for  $z \geq 0$  and  $\delta(z) = 0$  for  $z < 0$ .

From (1) and (2) one can readily derive equations for the corresponding expectations by taking the partial derivative of  $F(t; s)$  and  $F(t, x; s)$  with respect to  $s$  at the point  $s = 1$ . Using the notation:  $A(t) = E\{Z(t)\}$ ,  $A(t, x) = E\{Z(t, x)\}$ ,  $m = h'(1)$ , we obtain from (1) and (2)

$$A(t) = pm \int_0^t A(t-u) dG(u) + p[1 - G(t)] + (1-p)[1 - L(t)], \quad (3)$$

$$A(t, x) = pm \int_0^t A(t-u, x) dG(u) + \delta(x-t)\{p[1 - G(t)] + (1-p)[1 - L(t)]\}, \quad (4)$$

with the initial conditions:  $A(0) = 1 = A(0, 0)$ .

Equation (3) is a renewal type equation which has the following solution [15]:

$$A(t) = \sum_{k=0}^{\infty} (pm)^k [W * G^{*k}](t),$$

where  $W(t) = p[1 - G(t)] + (1-p)[1 - L(t)]$  and  $G^{*k}(t)$  is the  $k$ -fold convolution of  $G(t)$ . Note that  $G^{*1}(t) = G(t)$ ,  $G^{*k}(t) = \int_0^t G^{*(k-1)}(t-x) dG(x)$ , and  $[W * G^{*k}](t) = \int_0^t W(t-u) dG^{*k}(u)$ . The function  $H(t) = \sum_{k=0}^{\infty} G^{*k}(t)$  is termed the renewal function. The above solution of equation (3) assumes an explicit analytical form if  $G(t)$  is specified as a gamma distribution which is the most common choice in cell kinetics studies [15]. The shape of this distribution is quite flexible and has been supported by a vast body of experimental evidence [15].

Let  $\alpha$  be a real root of the characteristic equation

$$pm \int_0^{\infty} e^{-\alpha u} dG(u) = 1. \quad (5)$$

*Remark 1.* If  $M = pm > 1$  there exists a unique  $\alpha$  which is known to be positive (supercritical case). If  $M = 1$  then  $\alpha = 0$  (critical case). If  $M < 1$  then it is possible that the equation (5) has no solution, but if it exists it has to be negative (subcritical case).

Suppose that equation (5) has a unique solution  $\alpha$ . The parameter  $\alpha$  is a critical parameter (known also as the Malthusian parameter) that determines the asymptotic behavior of the process  $Z(t)$ . Now we need the notation:

$$A_{\alpha}(t) = A(t)e^{-\alpha t}, \quad A_{\alpha}(t, x) = A(t, x)e^{-\alpha t},$$

$$G_{\alpha,p}(t) = pm \int_0^t e^{-\alpha u} dG(u),$$

$$W_{\alpha,p}(t) = e^{-\alpha t}\{p[1 - G(t)] + (1-p)[1 - L(t)]\} \text{ and}$$

$$W_{\alpha,p,x}(t) = e^{-\alpha t}\delta(x-t)\{p[1 - G(t)] + (1-p)[1 - L(t)]\}.$$

One can see that (3) and (4) are equivalent to the following renewal-type equations

$$A_{\alpha}(t) = \int_0^t A_{\alpha}(t-u) dG_{\alpha,p}(u) + W_{\alpha,p}(t), \quad (6)$$

$$A_{\alpha}(t, x) = \int_0^t A_{\alpha}(t-u, x) dG_{\alpha,p}(u) + W_{\alpha,p,x}(t). \quad (7)$$

These equations have unique solutions which are bounded on every finite interval. The solutions are given by

$$A_\alpha(t) = \int_0^t W_{\alpha,p}(t-u) dH_{\alpha,p}(u) = \sum_{k=0}^{\infty} (W_{\alpha,p} * G_{\alpha,p}^{*k})(t), \quad (8)$$

$$A_\alpha(t, x) = \int_0^t W_{\alpha,p,x}(t-u) dH_{\alpha,p}(u) = \sum_{k=0}^{\infty} (W_{\alpha,p,x} * G_{\alpha,p}^{*k})(t), \quad (9)$$

where  $H_{\alpha,p}(t) = \sum_{k=0}^{\infty} G_{\alpha,p}^{*k}(t)$  is the renewal function and  $G_{\alpha,p}^{*k}(t)$  is the  $k$ -fold convolution of  $G_{\alpha,p}(t)$ .

*Remark 2.* In the most practically important case of binary splitting, the offspring p.g.f. is of the form:  $h(s) = 1 - \beta + \beta s^2$ , so that  $m = 2\beta$  has to be substituted for  $m$  in the characteristic equation (5).

### 3.2. A Renewal-type Non-homogeneous Immigration

Let  $Y(t)$  be the number of PCs at time  $t$  with the same evolution as that assumed for the process  $Z(t)$  (Section 3.1.1) but in the presence of immigration of new cells as described below. In like manner, we introduce the process  $Y(t, x)$  to represent the number of PCs of age  $\leq x$  at time  $t$  in the presence of immigration. Note that if  $x \geq t$  then  $Y(t) = Y(t, x)$ . Let  $\Psi(t; s) = E\{s^{Y(t)}\}$  and  $\Psi(t, x; s) = E\{s^{Y(t,x)}\}$  be the p.g.f.'s of the processes  $Y(t)$  and  $Y(t, x)$ , respectively. The process  $\bar{Y}_t(y)$  represents the number of PCs with the residual lifetime greater than  $y$  at time  $t$  (see Appendix 2).

The time periods  $\{T_i\}$  between the successive events of immigration (inter-arrival times) are independent and identically distributed (i.i.d.) r.v.'s with c.d.f.  $G_0(x) = \Pr(T_i \leq x)$ . At any given moment of immigration  $t > 0$ , the number of arrivals (immigrants) denoted by  $\nu(t)$  is random with distribution  $q_k(t) = \Pr\{\nu(t) = k\}$ ,  $\sum_{k=0}^{\infty} q_k(t) = 1$ . In other words, the moments of immigration form a renewal process of the form:

$$\mu(t) = \max \left\{ n : \sum_{i=1}^n T_i \leq t \right\}$$

but the corresponding counting process,  $I(t)$ , for the number of renewal events in the interval  $(0, t]$ , is non-homogeneous in time. If  $G_0(x) = 1 - e^{-rx}$ , the process  $\mu(t)$  reduces to a Poisson process with cumulative rate  $R(t) = rt$ .

Denote the p.g.f. of the number of immigrants at time  $t$  by  $g(t; s) = \sum_{k=0}^{\infty} q_k(t) s^k$  and let  $\gamma(t) = \frac{\partial g(t; s)}{\partial s} \Big|_{s=1}$  be the mean number of immigrants at time  $t$ . The the following integral equations hold

$$\Phi(t; s) = \int_0^t \Phi(t-u; s) g(t-u; F(t-u; s)) dG_0(u) + 1 - G_0(t), \quad (10)$$

$$\Phi(t, x; s) = \int_0^t \Phi(t-u, x; s)g(t-u; F(t-u, x; s))dG_0(u) + 1 - G_0(t), \quad (11)$$

with the initial conditions:  $\Phi(0; s) = \Phi(0, x; s) = 1$ , where  $F(t; s)$  and  $F(t, x; s)$  are defined by (1) and (2). Much like equations (1) and (2), these equations are derived by conditioning on the evolution of the first immigrant cell and applying the LTP.

From equations (10) and (11), we obtain equations for the corresponding expectations  $N(t) = E\{Y(t)\} = \partial\Phi(t; 1)/\partial s$ ,  $N(t, x) = E\{Y(t, x)\} = \partial\Phi(t, x; 1)/\partial s$ ,

$$N(t) = \int_0^t N(t-u)dG_0(u) + \int_0^t A(t-u)\gamma(t-u)dG_0(u), \quad (12)$$

$$N(t, x) = \int_0^t N(t-u, x)dG_0(u) + \int_0^t A(t-u, x)\gamma(t-u)dG_0(u), \quad (13)$$

with the initial conditions:  $N(0) = N(0, x) = 0$ .

Whenever  $A(t)$  is defined by (3) and the function  $V(t) = \int_0^t A(t-u)\gamma(t-u)dG_0(u)$  is bounded, equation (12) has a unique solution which is bounded on any finite interval. This solution has the form:

$$N(t) = \int_0^t V(t-u)dH(u) = \sum_{k=0}^{\infty} (V * G_0^{*k})(t), \quad (14)$$

where  $H(t) = \sum_{k=0}^{\infty} G_0^{*k}(t)$  is the corresponding renewal function.

Similarly, recalling the definition of  $A(t; x)$  given by (4) and assuming that the function  $V_x(t) = \int_0^t A(t-u, x)\gamma(t-u)dG_0(u)$  is bounded, we state that equation (13) has a unique solution  $N(t, x)$  which is bounded on any finite interval. The solution is given by

$$N(t, x) = \int_0^t V_x(t-u)dH(u) = \sum_{k=0}^{\infty} (V_x * G_0^{*k})(t). \quad (15)$$

The above results can be extended to include the case where the p.g.f.  $h(s)$  depends on the age of a cell (Sevastyanov's process), but such a generalization is of no particular value in biological applications because it is very difficult to specify this type of dependence on mechanistic grounds.

### 3.3. A Generalized Non-homogeneous Poisson Immigration

As discussed in Section 2.1, there is strong experimental evidence supporting the hypothesis of asymmetric division of stem cells in the CNS. It is also plausible that, at least under normal conditions, stem cells divide asymmetrically in all other renewing tissues. If we hypothesize in addition that the dynamics of the population of stem cells follows the model of Smith

and Martin [45], then the immigration process for the PC compartment is expected to be Poisson with time-dependent intensity  $r(t)$ . Indeed, the asymmetric mitosis of stem cells ensures ordinariness of the immigration process while the Smith and Martin model postulates a Markovian transition of resting stem cells to the state of active proliferation.

There is another sound reason for treating the immigration process as Poisson. We can consider the immigration of PCs as a superposition of many independent sparse point processes because many rare events of stem cell differentiation contribute to the influx of PCs at time  $t$ . The asymmetry of mitotic division of stem cells implies that only a single PC may arrive at each time point in the combined immigration process, thereby ensuring its ordinariness. Associated with each point process is a counting process  $\mu(t)$ ,  $t \geq 0$ , in continuous time, whose value at time  $t$  counts the number of events that have occurred in the interval  $(0, t]$ . Generally speaking, a counting process is an absolutely additive function defined on all Borel subsets of the real line which can take on only non-negative integer values. Requiring that each of the constituent counting processes be sparse (infinitesimally small), Grigelionis [46] proved that their superposition converges weakly to a Poisson counting process when the number of the constituent point processes increases. The limiting Poisson process can be non-homogeneous while the constituent processes do not need to be identical. While the exact formulation of the conditions used in Grigelionis' theorem is not important here, the result is general enough to expect the immigration of PCs to be approximately Poisson with rate  $r(t)$ . The first limiting result of a similar nature was obtained by Cox and Smith [47] for stationary renewal processes (see also [48, 49]).

Alternatively, one can hypothesize that the events of stem cell differentiation occur in the resting state. For each resting stem cell this is a rare event. Since the progenitor cell immigration process is formed by the superposition of such rare events, the result of Grigelionis suggests that the Poisson character of the immigration process is quite plausible under this hypothesis as well.

The above line of reasoning has to do with ordinary point processes allowing no tied events at any given time  $t$ . However, the Poisson immigration process can be generalized to allow for more than one arrival at time  $t$ . In this generalized process, the random arrival times are still governed by a non-homogeneous Poisson process with intensity  $r(t)$ , but the number of arrivals, given the event of immigration, may be random with an arbitrary p.g.f.  $g(s)$ . This model furnishes the possibility to include symmetric division of stem cells giving rise to two immigrant PCs that arrive in the PC population in pairs. More details on the model thus generalized is presented below.

Let  $Y(t)$  be the number of PCs at  $t$  with p.g.f.  $\Psi(t; s) = E\{s^{Y(t)}\}$ . For the process  $Y(t)$  we assume the same evolution as described in Section 3.1.1. Now we specify a model for the immigration process. Let  $0 = S_0 < S_1 < S_2 < S_3 < \dots$  be a sequence of time points in a non-

homogeneous Poisson process  $\xi(t)$  with rate  $r(s)$ . We use the notation  $R(t)$  for the cumulative rate  $R(t) = \int_0^t r(u)du$ . If  $T_i = S_i - S_{i-1}$  then  $S_k = \sum_{i=1}^k T_i$ ,  $k = 1, 2, \dots$ . Associated with every point  $S_k$  is an independent immigration component  $I_k$ , where  $\{I_k\}$  are i.i.d. r.v.'s with p.g.f.  $g(s) = Es^{I_k} = \sum_{i=0}^{\infty} q_i s^i$ .

The process  $Y(t)$  can be represented as

$$Y(t) = \sum_{k=1}^{\xi(t)} Z_{(k)}(t - S_k) \quad \text{if } \xi(t) > 0; \quad Y(t) = 0 \quad \text{if } \xi(t) = 0, \quad (16)$$

where  $Z_{(k)}(t)$  are i.i.d. branching processes with the same evolution as that of  $Z(t)$  but started with a random number of ancestors  $I_k$ . The p.g.f. of each  $Z_{(k)}(t)$  is given by  $F^*(t; s) = g(F(t; s))$ , where  $F(t; s)$  is a solution of the integral equation (1).

As shown in Appendix 1, the following important expression holds true:

$$\Psi(t; s) = \exp\left\{-\int_0^t r(t-u)[1 - F^*(u; s)]du\right\}. \quad (17)$$

Using (17) and (1) it is possible to calculate all moments of the process  $Y(t)$ . For illustration, let us derive the expected value  $EY(t) = \frac{\partial}{\partial s} \Psi(t; s)|_{s=1}$ . Introducing the expected number of immigrants  $\gamma = EI_k = g'(1)$ , we obtain

$$N(t) = \gamma \int_0^t r(t-u)A(u)du, \quad N(0) = 0, \quad (18)$$

where  $A(t)$  can be found from equation (3). After integration by parts in formula (18), we have

$$R(t) - \int_0^t R(t-u)dA(u) = N(t)/\gamma, \quad (19)$$

i.e., the equation for  $R$  assumes the form of a Volterra equation of the second kind which has a unique solution under mild regularity conditions [50]. This is the most basic fact in the theory under consideration that facilitates both computer simulation of the process  $Y(t)$  and estimation of model parameters in practical applications. We have more to say on this subject in Section 5.

*Remark 3.* If the immigration process is ordinary, we set  $\gamma = 1$  in equations (18) and (19). When allowing for the symmetric division of a stem cell with both daughter cells differentiating into PCs, we have to set  $\gamma = 2$ . If  $Z(t)$  is a Markov branching process with Malthusian parameter  $\alpha$ , then  $A(t) = e^{\alpha t}$  and it is not difficult to obtain the following solution of equation (19):

$$R(t) = N(t)/\gamma - \alpha \int_0^t N(x)dx. \quad (20)$$

Let us now consider the process  $Y(t, x)$  representing the number of PCs of age  $\leq x$  at time  $t$ . Note that if  $x \geq t$  then  $Y(t) = Y(t, x)$ . Proceeding from equation (2), one can obtain

the p.g.f.  $\Psi(t, x; s) = E\{s^{Y(t,x)} | Y(0, 0) = 0\}$  of this process under the model of generalized Poisson immigration. The derivation is similar to that for expression (17). The result is given by

$$\Psi(t, x; s) = \exp\left\{-\int_0^t r(t-u)[1 - F^*(u, x; s)]du\right\} \quad (21)$$

with the initial condition:  $\Psi(0, 0; s) = 1$ .

For the expectation  $N(t, x) = E\{Y(t, x)\} = \frac{\partial \Psi(t, x; s)}{\partial s} |_{s=1}$ , one can obtain the following expression

$$N(t, x) = \gamma \int_0^t r(t-u)A(u, x)du, \quad N(0, x) = 0, \quad (22)$$

where  $A(u, x)$  is a solution of the equation (4).

### 3.4. Homogeneous Poisson Immigration

For illustration purposes, we consider two examples of branching stochastic processes with homogeneous Poisson immigration. Their usefulness lies with the fact that more complex branching models can always be verified by reduction to these particular cases.

#### 3.4.1. The Bellman-Harris Process with Homogeneous Poisson Immigration

Setting  $p = 1$ , the process  $Z(t)$  reduces to the classical Bellman-Harris process with the cell evolution determined by the functions  $G(t)$  and  $h(s)$ . In what follows we will refer to this model structure as the  $(G, h)$ -evolution. We assume in addition that the immigration of PCs is a homogeneous Poisson process with parameter  $r$  and the p.g.f. of the number of immigrants is given by  $g(s) = \sum_{i=0}^{\infty} q_i s^i$ . Then formula (17) simplifies to

$$\Psi(t; s) = \exp\left\{-r \int_0^t [1 - F^*(u; s)]du\right\}, \quad (23)$$

where  $F^*(u; s) = g(F(t; s))$  and  $F(t; s)$  satisfies equation (10). Similarly from (11) and (21) one has

$$\Psi(t, x; s) = \exp\left\{-r \int_0^t [1 - F^*(u, x; s)]du\right\}. \quad (24)$$

Therefore by (14) and (18) (or directly from (23)) it follows

$$N(t) = \gamma r \int_0^t A(u)du, \quad N(0) = 0. \quad (25)$$

Similarly, by (15) and (22) (or directly from (24)) one obtains

$$N(t, x) = \gamma r \int_0^t A(u, x)du, \quad N(0, x) = 0. \quad (26)$$

Some asymptotic properties of this model can be readily ascertained. First of all, it follows from the well known asymptotic results (see, e.g., [7]) that

$$A(t) \sim A_0 e^{\alpha t}, \quad t \rightarrow \infty, \quad (27)$$

where

$$A_0 = A_0(m, \alpha) = \frac{\int_0^\infty e^{-\alpha u} (1 - G(u)) du}{m \int_0^\infty u e^{-\alpha u} dG(u)}, \quad (28)$$

and  $\alpha$  is the Malthusian parameter (determined by equation (5) with  $p = 1$ ).

In the subcritical case:  $\alpha < 0$  ( $m < 1$ ), it is assumed that

$$\int_0^\infty u e^{-\alpha u} dG(u) < \infty. \quad (29)$$

The situation is similar in the critical case  $\alpha = 0$  ( $m = 1$ ):

$$\int_0^\infty u dG(u) < \infty. \quad (30)$$

Note that in the supercritical case  $\alpha > 0$  ( $m > 1$ ), the condition (29) is automatically met.

Let us now consider equation (7) with  $p = 1$ . Under conditions (29) and (30) we can apply the Key Renewal Theorem (see, e.g., [48]) to obtain the following asymptotic result:

$$A(t, x) \sim A(x) e^{\alpha t}, \quad t \rightarrow \infty, \quad (31)$$

where  $\alpha$  is the Malthusian parameter and

$$A(x) = A_{m, \alpha}(x) = \frac{\int_0^x e^{-\alpha u} (1 - G(u)) du}{m \int_0^\infty u e^{-\alpha u} dG(u)}. \quad (32)$$

Hence, under conditions (29) and (30), it follows from (25) and (26) that the following asymptotic behavior holds true as  $t \rightarrow \infty$ :

(i) If  $\alpha < 0$  then

$$N(t) \rightarrow -\gamma r A_0 / \alpha; \quad N(t, x) \rightarrow -\gamma r A(x) / \alpha; \quad (33)$$

(ii) If  $\alpha = 0$  then

$$N(t) \sim \gamma r t; \quad N(t, x) \sim \gamma r A(x) t; \quad (34)$$

(iii) If  $\alpha > 0$  then

$$N(t) \sim \gamma r A_0 e^{\alpha t} / \alpha; \quad N(t, x) \sim \gamma r A(x) e^{\alpha t} / \alpha. \quad (34)$$

### 3.4.2. The Markov Branching Process with Homogeneous Poisson Immigration.

In this case we assume

$$G(t) = 1 - e^{-\lambda t}, \quad t \geq 0. \quad (36)$$

Then from (5) it follows that  $\alpha = \lambda(mp - 1)$ . Now by (6) and (7) we obtain

$$\begin{aligned} A(t) &= \sum_{n=0}^{\infty} (mp)^n (f * \Gamma_{n,\lambda})(t), \\ A(t, x) &= \sum_{n=0}^{\infty} (mp)^n (f_x * \Gamma_{n,\lambda})(t), \end{aligned} \quad (38)$$

where

$$f(t) = pe^{-\lambda t} + (1-p)(1-L(t)), \quad f_x(t) = \delta(x-t)\{pe^{-\lambda t} + (1-p)(1-L(t))\}$$

and

$$\Gamma_{n,\lambda}(t) = \frac{\lambda^n}{(n-1)!} \int_0^t u^{n-1} e^{-\lambda u} du.$$

Assuming in addition that  $p = 1$  we arrive at the classical Markov branching process with

$$f(t) = e^{-\lambda t}, \quad f_x(t) = \delta(x-t)e^{-\lambda t}.$$

It follows from (37) and (38) that

$$A(t) = e^{\lambda(m-1)t} \quad \text{if } m \neq 1; \quad A(t) \equiv 1, \quad \text{if } m = 1 \quad (39)$$

$$A(t, x) = e^{\lambda(m-1)t}(1 - e^{-m\lambda x}) \quad \text{if } m \neq 1; \quad A(t, x) = 1 - e^{-\lambda x}, \quad \text{if } m = 1. \quad (40)$$

Therefore,

$$A(t, x) = A(t)(1 - e^{-m\lambda x}). \quad (41)$$

On the other hand, it follows from (25) and (39) that

$$N(t) = \frac{\gamma r (e^{\lambda(m-1)t} - 1)}{\lambda(m-1)}, \quad \text{if } m \neq 1; \quad N(t) = \gamma r t \quad \text{if } m = 1. \quad (42)$$

Similarly, by (26) and (40) one can prove that

$$N(t, x) = \frac{\gamma r (e^{\lambda(m-1)t} - 1)(1 - e^{-m\lambda x})}{\lambda(m-1)} \quad \text{if } m \neq 1; \quad N(t, x) = \gamma r t (1 - e^{-\lambda x}) \quad \text{if } m = 1. \quad (43)$$

From (42) and (43) we finally obtain

$$N(t, x) = N(t)(1 - e^{-m\lambda x}). \quad (44)$$

### 3.5. A Two-type Age-dependent Branching Process with Immigration

We now intend to discuss a more general case of multi-type branching processes. More specifically, we show how one more cell type originating from PCs can be added to the model considered in Section 3.3. Suppose that the cells of this new type may divide or die but they are no longer capable of differentiating into another cell type. The resultant cell system consisting of two types of cells corresponds to the one described in the leukemia example of Section 2.2 and shown in Figure 3. The other example of oligodendrocyte development *in vivo* calls for an additional cell type (oligodendrocytes) resulting in a system with three types of cells (see Figure 2). Although the latter system is more complex, it can be modeled in a similar way. Let us describe the evolution of cells under the extended model with two types of cells.

We assume that every cell of type  $T_1$  with probability  $p$  has life-time  $\xi$  distributed in accordance with c.d.f.  $G(x) = \Pr(\xi \leq x)$  and, at the end of its life (mitotic cycle), it gives rise to  $\nu$  offsprings with p.g.f.  $h(s) = Es^\nu = \sum_{k=0}^{\infty} p_k s^k$ ,  $h(1) = 1$ . With probability  $1 - p$  every cell of type  $T_1$  differentiates into a cell of type  $T_2$ . The time to differentiation  $\eta$  has c.d.f.  $L(x) = P(\eta \leq x)$ . Every cell of type  $T_2$  has life-time  $\zeta$  with c.d.f.  $K(x) = P(\zeta \leq x)$  and, at the end of its life, it gives rise to  $\chi$  offsprings with p.g.f.  $\varphi(z) = Ez^\chi = \sum_{k=0}^{\infty} b_k z^k$ ,  $\varphi(1) = 1$ .

The above model allows for multiple outcomes of the mitotic cycle. In biological applications, the p.d.f.  $h(s)$  should be reduced to a simpler form:  $h(s) = 1 - \beta + \beta s^2$ , implying that a cell of type  $T_1$  divides with probability  $\beta$  or dies with probability  $1 - \beta$ . In like manner, for cells of type  $T_2$  we can assume that  $\varphi(s) = 1 - b + bs^2$ , implying the event of division with probability  $b$  and the event of death with probability  $1 - b$ . However, it makes sense to discuss a more general case with arbitrary p.g.f.'s  $h(s)$  and  $\varphi(s)$ .

Let  $Z(t)$  denote the number of  $T_1$  cells and  $X(t)$  the number of  $T_2$  cells at time  $t$ . Introduce the joint p.g.f.

$$F_1(t; s, z) = E\{s^{Z(t)} z^{X(t)} | Z(0) = 1\}$$

for  $Z(t)$  and  $X(t)$  given the process starts with a single  $T_1$  cell of zero age. Note that the marginal p.g.f.  $F_1(t; s, 1) = F(t; s) = E\{s^{Z(t)} | Z(0) = 1\}$  satisfies equation (1). Let  $F_2(t; z) = E\{z^{X(t)} | X(0) = 1\}$  be the p.g.f. of the number of  $T_2$  cells at time  $t$  given the process starts with one  $T_2$  cell of zero age.

Then the following system of integral equations holds for the p.g.f.'s  $F_1(t; s, z)$  and  $F_2(t; z)$ :

$$\begin{aligned} F_1(t; s, z) &= p \int_0^t h(F_1(t-u; s, z)) dG(u) + (1-p) \int_0^t F_2(t-u; z) dL(u) \\ &+ s\{p[1-G(t)] + (1-p)[1-L(t)]\} + z(1-p)L(t), \end{aligned} \quad (45)$$

$$F_2(t; z) = \int_0^t \varphi(F_2(t-u; z))dK(u) + z(1 - K(t)), \quad (46)$$

with the initial conditions:  $F_1(0; s, z) = s$  and  $F_2(0; z) = z$ .

Introduce the moments

$$A(t) = E\{Z(t)|Z(0) = 1\} = \frac{\partial F_1(t; s, z)}{\partial s} \Big|_{s=z=1} = \frac{\partial F(t; s)}{\partial s} \Big|_{s=1},$$

$$B(t) = E\{X(t)|Z(0) = 1\} = \frac{\partial F_1(t; s, z)}{\partial z} \Big|_{s=z=1},$$

$$C(t) = E\{X(t)|X(0) = 1\} = \frac{\partial F_2(t; z)}{\partial z} \Big|_{z=1}, \quad c = \varphi'(1).$$

Note that  $A(t)$  can be obtained from equation (3). From (45) and (46), we also obtain equations for  $B(t)$  and  $C(t)$ :

$$B(t) = pm \int_0^t B(t-u)dG(u) + (1-p) \int_0^t C(t-u)dL(u) + (1-p)L(t), \quad (47)$$

$$C(t) = c \int_0^t C(t-u)dK(u) + 1 - K(t), \quad (48)$$

with the initial conditions:  $B(0) = 0, C(0) = 1$ . These equations have the same structure as (3) and can be solved by the same method.

Let  $Z^I(t)$  be the number of  $T_1$  cells and  $X^I(t)$  be the number of  $T_2$  cells in the presence of non-homogeneous Poisson immigration. We introduce the notation:  $\Psi^I(t; s, z) = E\{s^{Z^I(t)}z^{X^I(t)}|Z^I(0) = X^I(0) = 0\}$ . Then by the same argument as that given in Appendix 2 one can prove that

$$\Psi^I(t; s, z) = \exp\left\{-\int_0^t r(t-u)[1 - g(F_1(u; s, z))]du\right\} \quad (49)$$

with the initial condition:  $\Psi^I(0; s, z) = 1$ . Here the p.g.f.  $F_1(t; s, z)$  is defined by equation (45) and  $g(s)$  is the p.g.f. of the number of immigrants. Therefore all moments can be exactly computed from formula (49). In particular, the mean values are given by

$$N^I(t) = EZ^I(t) = \frac{\partial \Psi^I(t; s, z)}{\partial s} \Big|_{s=z=1} = \int_0^t r(t-u)A(u)du$$

and

$$M^I(t) = \frac{\partial \Psi^I(t; s, z)}{\partial z} \Big|_{s=z=1} = \int_0^t r(t-u)B(u)du.$$

### 3.6. The Case of Initial Random Number of Cells

When analyzing actual data on renewing cell populations *in vivo*, it is important to accommodate the case where the process starts with random numbers of cells in each compartment. For example, consider the model of Section 3.5 assuming now that every process starts with random numbers  $\xi_1$  and  $\xi_2$  of  $T_1$  and  $T_2$  cells, respectively. The random vector  $(\xi_1, \xi_2)$  has a p.g.f.

$g(s_1, s_2) = E(s_1^{\xi_1} s_2^{\xi_2})$ . We also need the notation:  $g_1(s_1) = g(s_1, 1) = E(s_1^{\xi_1})$  and  $g_2(s_2) = g(1, s_2) = E(s_2^{\xi_2})$ . In particular, if  $\xi_1$  and  $\xi_2$  are independent then  $g(s_1, s_2) = g_1(s_1)g_2(s_2)$ . Denote  $m_1 = E\xi_1 = g'_1(1)$ ,  $m_2 = E\xi_2 = g'_2(1)$ . Let  $Z^*(t)$  and  $X^*(t)$  be the number of  $T_1$  and  $T_2$  cells, respectively, produced by  $\xi_1$  cells of type  $T_1$ . Then we have

$$F_1^*(t; s, z) = E\{s^{Z^*(t)} z^{X^*(t)} | Z^*(0) = \xi_1\} = g_1(F_1(t; s, z)). \quad (50)$$

Let  $X^-(t)$  be the number of  $T_2$  cells produced by  $\xi_2$  cells of type  $T_2$ . Then

$$F_2^-(t; z) = E\{z^{X^-(t)} | X^-(0) = \xi_2\} = g_2(F_2(t; z)). \quad (51)$$

Finally, let  $Y_k^+(t)$ ,  $k = 1, 2$ , be the number of type  $T_k$  cells at time  $t$  in the two-type branching process under consideration given it starts with  $\xi_1$  cells of type  $T_1$  and  $\xi_2$  cells of type  $T_2$ . Then

$$Y_1^+(t) = Z^I(t) + Z^*(t), \quad Y_2^+(t) = X^I(t) + X^*(t) + X^-(t), \quad (52)$$

where the three vectors  $(Z^I(t), X^I(t))$ ,  $(Z^*(t), X^*(t))$  and  $X^-(t)$  are independent.

Therefore, the corresponding joint p.g.f. is given by

$$\begin{aligned} \Psi^+(t; s, z) &= E\{s^{Y_1^+(t)} z^{Y_2^+(t)} | Y_1^+(0) = \xi_1, Y_2^+(0) = \xi_2\} = \Psi^I(t; s, z) F_1^*(t; s, z) F_2^-(t; z) \\ &= \Psi^I(t; s, z) g_1(F_1(t; s, z)) g_2(F_2(t; z)), \end{aligned} \quad (53)$$

where the p.g.f.  $\Psi^I(t; s_1, s_2)$ ,  $F_1(t; s_1, s_2)$  and  $F_2(t; s_2)$  are defined by equations (45), (46) and (49) of Section 3.5. Now it is possible to derive formulas for all moments. In particular, it follows from (50) and (51) that

$$A^*(t) = EZ^*(t) = m_1 A(t), \quad B^*(t) = EX^*(t) = m_1 B(t), \quad B^-(t) = EZ_2^-(t) = m_2 C(t). \quad (54)$$

Similarly, from (52), (53) and (54) we obtain the expected values

$$M_1^+(t) = EY_1^+(t) = N^I(t) + m_1 A(t), \quad M_2^+(t) = EY_2^+(t) = M^I(t) + m_1 B(t) + m_2 C(t). \quad (55)$$

To derive the variance for each cell type and the covariance between the two types of cells, we proceed as follows:

$$\begin{aligned} \text{Var} Z^*(t) &= m_1 \text{Var} Z(t) + d_1^2 A^2(t), \\ \text{Var} X^*(t) &= m_1 \text{Var} X(t) + d_1^2 B^2(t), \\ \text{Var} X^-(t) &= m_2 \text{Var}\{X(t) | X(0) = 1\} + d_2^2 C^2(t). \\ E Z^*(t) X^*(t) &= m_1 E Z(t) X(t) + v_1 A(t) B(t), \end{aligned}$$

$$\text{Cov}\{Z^*(t), X^*(t)\} = m_1 \text{Cov}\{Z(t), X(t)\} + d_1^2 A(t)B(t).$$

From these formulas together with (50)-(55) we finally obtain

$$\begin{aligned} \text{Var}Y_1^+ &= \text{Var}Z^I(t) + m_1 \text{Var}Z(t) + d_1^2 A^2(t), \\ \text{Var}Y_2^+ &= \text{Var}X^I(t) + m_1 \text{Var}Z(t) + d_1^2 A^2(t) + m_2 \text{Var}\{X(t)|X(0) = 1\} + d_2^2 C^2(t), \\ \text{Cov}(Y_1^+(t), Y_2^+(t)) &= \text{Cov}(Y_1(t), Y_2(t)) + m_1 \text{Cov}\{Z_1(t), Z_2(t)\} + d_1^2 A(t)B(t). \end{aligned}$$

## 4. Analysis of Double Labeling Experiments.

Due to their complexity, proving identifiability of the preceding models is a very difficult task. However, it is entirely possible that the ultimate stochastic model would have been nonidentifiable if the numbers (or fractions) of cells of different types were the only experimental outcomes accessible to measurement. This obstacle can be surmounted by independently estimating the mitotic cycle parameters from specially designed labeling experiments. There are two exogenous markers: BrdU and Ki-67 that can be used for this purpose. While BrdU represents a relatively stable marker for labeling the *S*-phase *in vivo*, Ki-67 is a nuclear protein expressed in all periods of the cell cycle except the resting phase [51]. Both proliferative markers have been studied extensively in the context of neurogenesis [51]. Using Ki-67 it is possible to label all proliferating cells in *in vitro* samples of the (dissociated) spinal cord obtained from experimental animals at different times after a pulse BrdU label administered *in vivo* at time  $t = 0$ . Leukemic cells are even easier to handle with this technique. Analyzing the kinetics of cells that have been pulse-labeled with BrdU on a fluorescence activated cell sorter is a routine experimental technique for studying the developmental fate of dividing precursor cells.

Let  $\Lambda(t)$  be an age-dependent branching process representing the number of doubly labeled cells at time  $t > 0$ , and let  $I(0)$  be the number of cells in the *S*-phase (labeled with BrdU) at time  $t = 0$ . Suppose that, upon the completion of its MC, every PC divides with probability  $\beta$  or exits the cycle with probability  $1 - \beta$ . This gives the offspring p.g.f. of the form:  $h(s) = 1 - \beta + \beta s^2$ . We also assume that the lengths of the MC phases are independent r.v.'s. The c.d.f.'s  $F_S$  and  $F_{G_2+M}$  stand for the *S* and  $G_2 + M$  phases, respectively. The distribution of the whole MC duration is denoted by  $G$ . The same model structure may be adopted for different subtypes of progenitor cells.

Let  $Z(t)$  be a Bellman-Harris branching process (without immigration) representing the number of cells produced by a single cell with the above-described  $(G, h)$ -evolution and denote its p.g.f. by  $\psi(t; s) = E\{s^{Z(t)}|Z(0) = 1\}$ . The p.g.f.  $\varphi(t; s) = E\{s^{\Lambda(t)}|\Lambda(0) = 1\}$  satisfies the integral equation

$$\varphi(t; s) = \int_0^t h[\psi(t-u; s)] dW_0(u) + s[1 - W_0(t)], \quad \varphi(0, s) = s \quad |s| \leq 1 \quad (56)$$

$$\psi(t; s) = \int_0^t h[\psi(t-u; s)]dG(u) + s[1 - G(t)], \quad \psi(0, s) = s \quad |s| \leq 1 \quad (57)$$

where  $W_0 = \bar{U}_S(t, 0) * F_{G_2+M}(t)$ , and  $\bar{U}_S(t, 0)$  is the stationary c.d.f. of the residual time in the  $S$ -phase given the start of the labeling experiment at time  $t = 0$  and the symbol  $*$  stands for the convolution. The c.d.f.  $\bar{U}_S(t, 0)$  is essentially a limiting distribution requiring an indefinitely long prehistory of the cell population (see Appendix 2).

If the total rate of entry of cells into the  $S$ -phase is constant in time (e.g., if the rate of immigration is constant and cells do not divide), the function  $\bar{U}_S(t, 0)$  has the form [15]:

$$\bar{U}_S(t, 0) = \frac{1}{\bar{\tau}_S} \int_0^t [1 - F_S(x)]dx, \quad (58)$$

where  $\bar{\tau}_S$  is the mean duration of the  $S$ -phase. From the practical standpoint, for validity of formula (58) it is only required that the rate of cell entry into the  $S$ -phase is approximately constant for a short (equaling, say, the 0.9th quantile of the distribution  $F_S$ ) period of time immediately preceding the administration of BrdU [15]. Another formula for  $\bar{U}_S(t, 0)$  can be derived in the supercritical (exponential growth) and critical cases in the absence of immigration. It has the form [15]:

$$\bar{U}_S(t, 0) = 1 - \frac{e^{\alpha t} \int_t^\infty e^{-\alpha x} [1 - F_S(x)]dx}{\int_0^\infty e^{-\alpha x} [1 - F_S(x)]dx}, \quad (59)$$

where  $\alpha \geq 0$  is the Malthusian parameter. It is a remarkable fact that formula (59) remains the same in the presence of a homogeneous Poisson immigration. In the subcritical case ( $\alpha < 0$ ) with a homogeneous Poisson immigration, this distribution is given by

$$\bar{U}_S(t, 0) = \frac{\int_0^t e^{-\alpha x} [1 - F_S(x)]dx}{\int_0^\infty e^{-\alpha x} [1 - F_S(x)]dx}. \quad (60)$$

The derivation of these surprising results is given in Appendix 2. From the practical standpoint, for formulas (59) and (60) to be valid it is only required that the immigration rate for PCs (and not necessarily the rate of entry in the MC) remains constant over a short (equaling, say, the 0.9th quantile of the  $S$ -phase distribution) period of time immediately preceding the administration of BrdU.

From equations (56) and (57), we obtain the following expressions for the mean  $M(t)$  and the variance  $D(t)$ :

$$M(t) = E\{\Lambda(t) | \Lambda(0) = 1\} = 1 - W_0(t) + \sum_{k=0}^{\infty} (2\beta)^{k+1} W_0 * [G^{*k} - G^{*(k+1)}(t)], \quad (61)$$

$$D(t) = \text{Var}\{\Lambda(t) | \Lambda(0) = 1\} = Q(t) + \sum_{k=0}^{\infty} (2\beta)^{k+1} (Q * G^{*(k+1)})(t) + M(t) - M^2(t), \quad (62)$$

where  $Q(t) = 2\beta(B^2 * W_0)(t)$ ,  $B(t) = \sum_{k=0}^{\infty} (2\beta)^{k+1} [G^{*k} - G^{*(k+1)}(t)]$ , and  $G^{*k}$  denotes the  $k$ -fold convolution of  $G$ . In the case of gamma-distributed phases of the MC the above expressions for  $M(t)$  and  $D(t)$  can be represented in closed form. To allow for inter-animal variability, we treat the initial number of cells  $I(0)$  in the  $S$ -phase as a r.v. with mean  $a_I$  and variance  $\sigma_I$ . Then formulas (61) and (62) are modified as follows (see Section 3.6):  $M_I(t) = a_I M(t)$ ,  $D_I(t) = a_I D(t) + \sigma_I^2 M^2(t)$ .

In the above formulas, the time of labeling was set at  $t = 0$ . By changing variables:  $t' = t - t_0$  we can set it at any time point  $t_0$ . The expression for  $M(t)$  and  $D(t)$  can be used for estimation of model parameters from the counts of doubly labeled cells:  $X^{obs}(t; t_0), t > t_0$ , in conjunction with the generalized method of moments [52, 53].

## 5. Some Remarks Pertaining to Future Applications

It is extremely difficult, if not impossible, to prove identifiability of the models discussed in this paper. A mathematical or simulation model whose parameters are not identifiable is of no use for data analysis, unless a proper reparametrization results in identifiable combinations of model parameters. If such combinations cannot be found, additional sources of information need to be utilized in order to overcome this difficulty. This is the reason why a two-step procedure described in the present paper deserves serious attention. At the first step of this procedure, a subset of model parameters can be estimated from specially designed labeling experiments. This subset typically includes such parameters as the probability of division, mean duration of the MC and the corresponding standard deviation. Having these parameters fixed and fitting the model to empirical moments of the counts of cells of different types as functions of time we can estimate the expected influx of PCs, characterized by the rate  $R(t)$ , and all the as yet unknown parameters. This two-step analysis is designed to avoid the problem of over-parametrization by keeping the number of unknown parameters to a minimum at each step. In reference to the example of Section 2.1, there will be only 7 parameters (3 for GRP cells, 3 for O-2A/OPCs, and 1 for oligodendrocytes) to be estimated from observed counts of the three cell types if one uses gamma-distributed life-times for all the cell transformations as a suitable approximation to reality. There will be only 5 such parameters (3 for LPCs and 2 for LBs) in the leukemia example of Section 2.2.

One distinct advantage of the proposed approach is worthy of notice. One needs to specify the immigration rate  $r(t)$  (or  $R(t)$ ) to model the dynamics of multi-type cell populations and design associated estimation procedures. Recalling equation (19), we see that this rate is completely determined by the mean number of PCs, given the values of other parameters are

fixed. For example, in the Markov case with  $\gamma = 1$  it follows from formula (20) that

$$r(t) = N'(t) - \alpha N(t) \tag{63}$$

where  $a$  is the Malthusian parameter. On the other hand, the rate  $r(t)$  can be treated as a simple one-parameter family given  $N(t)$  is known. Indeed, both  $N(t)$  and  $N'(t)$  can be estimated (using, e.g., kernel smoothing techniques) from observed cells counts and then substituted into equation (63). This means that the rate  $r(t)$  is known up to the parameter  $\alpha$  which can be estimated jointly with other parameters of PCs from higher moments (e.g., the variance) as functions of time. In the Bellman-Harris case, one can use an analytical solution of equation (19) to attain the same ends. Smoothing techniques are also warranted here in order to obtain a regularized solution [54]. By this simple expedient of transforming the mean number of PCs, we dramatically reduce the number of free parameters when specifying the unknown immigration rate. This idea is especially useful when resorting to simulations in order to facilitate the procedure of model fitting in complicated cases. With this approach, the most appropriate estimation technique is the method of moments or its generalized version [52, 53].

It should be noted that the methodology routinely used to analyze cell kinetics *in vivo* is overly simplistic. For example, the cumulative labeling methodology [55, 56] used to estimate the MC parameters *in vivo* is based on assumptions that can hardly be perceived as realistic. In particular, this method assumes a steady state of the cell population over an extended period of time (see our comment following formula (58)) and does not allow for variability of the mitotic cycle duration between cells.

The methodology proposed in the present communication represents a much more sophisticated but tractable alternative. The development of more sophisticated models and methods, however, can not serve as an end in itself. For a model to be useful in biological applications, its complexity must be adequate for the information contained in the data to be analyzed. From this perspective, the most important requirement is that the model be identifiable. Wherever possible, a theoretical proof of model identifiability should be provided. Numerical or simulation studies can also be conducted to show that the model is not over-parameterized and its parameters can be estimated from the data if sufficient sample size is provided. Alternatively, one needs additional sources of information to surmount the difficulty, and this is what we resorted to in the present paper. As long as the requirement of identifiability is met, the quest for generality is always warranted in model building. We believe that this general principle is more appropriate for the development of mechanistically motivated models of biological systems than the well-known principle of parsimony, especially when such models are intended for predictions and future analyses of similar data. The approach presented here appears to satisfy both conflicting requirements, thereby representing the desired compromise between identifiability and flexibility of the proposed modeling techniques. However, considerable room

remains for further exploration of computational and statistical aspects of their application to real data analysis.

### **Acknowledgments**

This research is supported by NIH/NINDS grant RO1 NS39511 (Yakovlev) and Grant MM-1101/02 (Yanev) from the NSF of the Bulgarian Ministry of Education and Science. We are very grateful to Drs. Mark Noble and Craig Jordan (University of Rochester) who helped us understand and describe the biological examples presented in this paper.

### **References**

- [1] I.J. Bienaymé. De la loi de la multiplication et de la durée des familles. Soc.Philomath. Paris Extraits Ser. 5 (1845) 37-39.
- [2] F. Galton, H.W. Watson, On the probability of the extinction of the families, J. Roy. Antropol. Soc. London 4 (1874) 138-144.
- [3] A.N. Kolmogorov, On the solution of a biological problem. Proc.of Tomsk University, 2 (1938) 7-12 (in Russian).
- [4] R. Bellman, T. Harris, On the theory of age-dependent stochastic branching processes, Proc. Natl. Acad. Sci. U.S.A. 34 (1948) 601-604.
- [5] R. Bellman, T. Harris, On age-dependent binary branching processes, AM 55 (1952) 280-295.
- [6] B.A. Sevastyanov, Limit theorem for branching stochastic processes of special form, TPA 2 (1957) 121-136 (in Russian).
- [7] T. Harris, The Theory of Branching Processes, Springer, Berlin, 1963.
- [8] B.A. Sevastyanov, Branching Processes, Nauka, Moscow, 1971 (in Russian).
- [9] C.J. Mode, Multitype Branching Processes, Elsevier, NY, 1971.
- [10] K.B. Athrya, P.E. Ney, Branching Processes, Springer, Berlin, 1972.
- [11] P. Jagers, Branching Processes with Biological Applications, Wiley, London, 1975.
- [12] S. Assmussen, H. Hering, Branching Processes, Birkhauser, Boston, 1983.

- [13] P. Gutterp, *Statistical Inference for Branching Processes*, Wiley, NY, 1991.
- [14] M. Kimmel, D.E. Axelrod, *Branching Processes in Biology*, Springer, NY, 2002.
- [15] A.Yu. Yakovlev, N. M. Yanev, *Transient Processes in Cell Proliferation Kinetics*, Springer, Berlin, 1989.
- [16] M.C. Raff, R.H. Miller, M. Noble, A glial progenitor cell that develops in vitro into an astrocyte or an oligodendrocyte depending on the culture medium, *Nature* 303(5916) (1983) 390-396.
- [17] A. Espinosa de los Monteros, M. Zhang, et al., O2A progenitor cells transplanted into the neonatal rat brain develop into oligodendrocytes but not astrocytes, *Proc. Natl. Acad. Sci. U.S.A.* 90(1) (1993) 50-54.
- [18] S. C. Barnett, A. M. Hutchins, M. Noble, et al., Purification of olfactory nerve ensheathing cells from the olfactory bulb, *Dev. Biol.* 155 (1993) 337-350.
- [19] N. Scolding, P. Rayner, et al., Identification of A2B5-positive putative oligodendrocyte progenitor cells and A2B5-positive astrocytes in adult human white matter, *Neuroscience* 89 (1999) 1-4.
- [20] M. Rao, M. Noble, M. Mayer-Proschel, A tripotential glial precursor cell is present in the developing spinal cord, *Proc. Natl. Acad. Sci. U.S.A.* 95 (1998) 3996-4001.
- [21] N. Gregori, C. Proschel, M. Noble, M. Mayer-Proschel, The tripotential glial-restricted precursor (GRP) cell and glial development in the spinal cord: Generation of bipotential oligodendrocyte-type-2 astrocyte progenitor cells and dorsal-ventral differences in GRP cell function, *J. Neurosci.* 22 (2002) 248-256.
- [22] M. Rao, M. Mayer-Proschel, Glial restricted precursors are derived from multipotent neuroepithelial stem cells, *Dev. Biology* 188 (1997) 48-63.
- [23] M. Mayer-Proschel, A. Kalyani, T. Mujtaba, M.S. Rao, Isolation of lineage-restricted neuronal precursors from multipotent neuroepithelial stem cells, *Neuron* 19 (1997) 773-785.
- [24] N. Hawkins, G. Garriga, Asymmetric cell division: from A to Z, *Genes and Development* 12 (1998) 3625-3638.
- [25] M.C. Mione, J.F.R. Cavanagh, B. Harris, J.G. Parnavelas, Cell fate specification and symmetrical/asymmetrical divisions in the developing cerebral cortex, *J. Neurosci.* 17 (1997) 2018-2029.

- [26] X. Qian, S.K. Goderie, Q. Shen, J.H. Stern, S. Temple, Intrinsic programs of patterned cell lineages in isolated vertebrate CNS ventricular zone cells, *Development* 125 (1998) 3143-3152.
- [27] X. Qian, Q. Shen, S.K. Goderie, W. He, A. Capela, A.A. Davis, S. Temple, Timing of CNS cell generation: A programmed sequence of neuron and glial cell production from isolated murine cortical stem cells, *Neuron* 28 (2000) 69-80.
- [28] C.B. Reid, S.F. Tavazoie, C.A. Walsh, Clonal dispersion and evidence for asymmetric cell division in ferret cortex, *Development* 124 (1997) 2441-2450.
- [29] Q. Shen, X. Qian, A. Capela, S. Temple, Stem cells in the embryonic cerebral cortex: their role in histogenesis and patterning, *J. Neurobiol.*, 36 (1998) 162-174.
- [30] Q. Shen, W. Zhong, Y.N. Jan, S. Temple, Asymmetric numb distribution is critical for asymmetric cell division of mouse cerebral cortical stem cells and neuroblasts, *Development* 129 (2002) 4843-4853.
- [31] K. Boucher, A.Y. Yakovlev, M. Mayer-Proschel, and M. Noble, A stochastic model of temporally regulated generation of oligodendrocytes in vitro, *Mathematical Biosciences* 159 (1999) 47-78.
- [32] K. Boucher, A.V. Zorin, A.Y. Yakovlev, M. Mayer-Proschel, M. Noble, An alternative stochastic model of generation of oligodendrocytes in cell culture, *J. Mathematical Biology* 43 (2001) 22-36.
- [33] E. von Collani, A. Tsodikov, A. Yakovlev, M. Mayer-Proschel, M. Noble, A random walk model of oligodendrocyte generation in vitro and associated estimation problems, *Mathematical Biosciences* 159 (1999) 189-204.
- [34] O. Hyrien, M. Mayer-Proschel, M. Noble, A.Y. Yakovlev, Estimating the life-span of oligodendrocytes from clonal data on their development in cell culture, *Mathematical Biosciences* 193 (2005) 255-274.
- [35] O. Hyrien, M. Mayer-Proschel, M. Noble, A.Y. Yakovlev, A stochastic model to analyze clonal data on multi-type cell populations, *Biometrics* 61 (2005) 199-207.
- [36] A.Y. Yakovlev, K. Boucher, M. Mayer-Proschel, M. Noble, Quantitative insight into proliferation and differentiation of oligodendrocyte type 2 astrocyte progenitor cells in vitro, *Proc. Natl. Acad. Sci. USA* 95 (1998) 14164-14167.
- [37] A.Yu. Yakovlev, M. Mayer-Proschel, M. Noble, A stochastic model of brain cell differentiation in tissue culture, *J. Mathematical Biology* 37 (1998) 49-60.

- [38] A. Yakovlev, E. von Collani, M. Mayer-Proschel, M. Noble, Stochastic formulations of a clock model for temporally regulated generation of oligodendrocytes in vitro, *Mathematical and Computer Modelling* 32 (2000) 125-137.
- [39] A.V. Zorin, A.Y. Yakovlev, M. Mayer-Proschel, M. Noble, Estimation problems associated with stochastic modeling of proliferation and differentiation of O-2A progenitor cells in vitro, *Mathematical Biosciences* 67 (2000) 109-121.
- [40] S. Temple, M.C. Raff, Clonal analysis of oligodendrocyte development in culture: Evidence for a developmental clock that counts cell division, *Cell* 44 (1986) 773-779.
- [41] C. T. Jordan, Cancer stem cell biology: from leukemia to solid tumors, *Curr. Opin. Cell Biol.* 16(6) (2004) 708-12.
- [42] R. Pardal, M.F. Clarke, et al., Applying the principles of stem-cell biology to cancer, *Nat. Rev. Cancer* 3(12) (2003) 895-902.
- [43] T. Reya, S.J. Morrison, et al., Stem cells, cancer, and cancer stem cells, *Nature* 414(6859) (2001) 105-11.
- [44] C.S. Lee, G. L. Yang, Multitype decomposable age-dependent branching processes with applications, *J. Appl. Probab.* 32 (1995) 591-608.
- [45] J.A. Smith, L. Martin, Do cells cycle?, *Proc. Natl. Acad. Sci. USA* 70 (1973) 1263-1267.
- [46] B. Grigelionis, On the convergence of sums of stepwise random processes to a poisson process, *Theory of Probability and its Applications* 8(2) (1963) 189-194 (in Russian).
- [47] D.R. Cox, W.L. Smith, On the superposition of renewal processes, *Biometrika* 41(1/2) (1954) 91-99.
- [48] W. Feller, *An Introduction to Probability Theory and Its Application*, Second Edition, John Wiley & Sons, NY, 1971.
- [49] A.Y. Khinchin, *Works on Queueing Theory*, Physmatgiz, Moscow, 1963 (in Russian).
- [50] P. Linz, *Analytical and Numerical Methods for Volterra Equations*, Philadelphia, SIAM, 1985.
- [51] N. Kee, S. Sivalingam, R. Boonstra, J.M. Wojtowicz, The utility of Ki-67 and BrdU as proliferative markers of adult neurogenesis, *J. Neurosci. Methods* 115 (2002) 97-105.

- [52] D.W.K. Andrews, B. Lu, Consistent model and moment selection procedures for GMM. Estimation with application to dynamic panel data models, *J of Econometrics* 101(1) (2001) 123-164.
- [53] L.P. Hansen, Large sample properties of generalized method of moments estimators, *Econometrica* 50(4) (1982) 1029-1054.
- [54] O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science*, **1**, 502-527.
- [55] T. Takahashi, R.S. Nowakowski, V.S. Caviness, Jr., The cell cycle of the pseudostratified ventricular epithelium of the embryonic murine cerebral wall, *J. Neurosci.* 15 (1995) 6046-6057.
- [56] R.S. Novakowski, V.S. Caviness Jr., T. Takahashi, N.L. Hayes, Population dynamics during cell proliferation and neurogenesis in the developing murine neocortex. In: *Cortical Development: From Specification to Differentiation*, Hohmann, C.F, ed., Springer, Berlin, 2002.

## Appendix 1. Derivation of the Probability Generating Function for the Age-dependent Branching Process with Non-homogeneous Poisson Immigration

Formula (17) gives the following basic expression for the p.g.f. of the process  $Y(t)$  introduced in Section 3.2:

$$\Psi(t; s) = \exp \left\{ - \int_0^t r(t-u) [1 - F^*(u; s)] du \right\},$$

where  $F^*(t; s) = g(F(t; s))$  and  $F(t; s)$  is a solution of the integral equation (1).

*Proof.* It follows from (16) that

$$\Psi(t; s) = E s^{Y(t)} = E \left\{ s^{\sum_{i=1}^{\xi(t)} Z_{(i)}(t-S_i)} \right\} = \sum_{n=0}^{\infty} P(\xi(t) = n) E \left\{ s^{\sum_{i=1}^{\xi(t)} Z_{(i)}(t-S_i)} \mid \xi(t) = n \right\}.$$

Now we have

$$\begin{aligned} \Delta_n(t; s) &= E \left\{ s^{\sum_{i=1}^{\xi(t)} Z_{(i)}(t-S_i)} \mid \xi(t) = n \right\} \\ &= \int_0^t \int_{u_1}^t \dots \int_{u_{n-1}}^t E \left\{ s^{\sum_{i=1}^n Z_{(i)}(t-u_i)} \right\} dP_{S_1, S_2, \dots, S_n}(u_1, u_2, \dots, u_n) \\ &= \int_0^t \int_{u_1}^t \dots \int_{u_{n-1}}^t \prod_{i=1}^n F^*(t-u_i; s) dP_{S_1, S_2, \dots, S_n}(u_1, u_2, \dots, u_n), \end{aligned}$$

where  $P_{S_1, S_2, \dots, S_n}(u_1, u_2, \dots, u_n) = \mathbf{P}(S_1 \leq u_1, S_2 \leq u_2, \dots, S_n \leq u_n)$ .

We will use the fact that  $\eta(t) = R(\xi(t))$  is a homogeneous Poisson process with rate  $\lambda = 1$  and  $R(S_k) = \Gamma_k$  has  $\Gamma(k, 1)$  distribution,  $k = 1, 2, \dots$ . Let  $R^{-1}(\cdot)$  be the inverse function of the rate  $R(t)$ . Then

$$\begin{aligned} P_{S_1, \dots, S_n}(u_1, u_2, \dots, u_n) &= \mathbf{P}\{R^{-1}(\Gamma_1) \leq u_1, \dots, R^{-1}(\Gamma_n) \leq u_n\} \\ &= P_{\Gamma_1, \dots, \Gamma_n}(R(u_1), R(u_2), \dots, R(u_n)). \end{aligned}$$

The Order Statistics Property yields

$$\begin{aligned} \Delta_n(t; s) &= \frac{n!}{R^n(t)} \int_0^t \int_{u_1}^t \dots \int_{u_{n-1}}^t \prod_{i=1}^n F^*(t - u_i; s) dR(u_n) \dots dR(u_2) dR(u_1) \\ &= (n!/R^n(t)) \int_0^t \int_{u_1}^t \dots \int_{u_{n-1}}^t \prod_{i=1}^n r(u_i) F^*(t - u_i; s) du_n \dots du_2 du_1 \\ &= (1/R^n(t)) \int_0^t \int_0^t \dots \int_0^t \prod_{i=1}^n r(u_i) F^*(t - u_i; s) du_1 du_2 \dots du_n \end{aligned}$$

by virtue of the fact that  $f(u_1, \dots, u_n) = \prod_{i=1}^n r(u_i) F^*(t - u_i; s)$  is a symmetric function.

Therefore we can write

$$\Delta_n(t; s) = (1/R^n(t)) \prod_{i=1}^n \int_0^t r(u_i) F^*(t - u_i; s) du_i = (1/R^n(t)) \left\{ \int_0^t r(u) F^*(t - u; s) du \right\}^n.$$

Then one has

$$\begin{aligned} \Psi(t; s) &= \sum_{n=0}^{\infty} P(\xi(t) = n) \Delta_n(t; s) \\ &= e^{-R(t)} \sum_{n=0}^{\infty} (R^n(t)/n!) (1/R^n(t)) \left\{ \int_0^t r(u) F^*(t - u; s) du \right\}^n \\ &= \exp\left\{-R(t) + \int_0^t r(u) F^*(t - u; s) du\right\} \\ &= \exp\left\{-\int_0^t r(u) [1 - F^*(t - u; s)] du\right\} \\ &= \exp\left\{-\int_0^t r(t - u) [1 - F^*(u; s)] du\right\} \end{aligned}$$

which proves (17).

## Appendix 2: Residual Life-time Distributions

For the processes  $Z(t)$  and  $Y(t)$ , we define the residual time distributions at time  $t$  as follows:

$$\begin{aligned} U_t(y) &= 1 - \bar{U}_t(y), \quad \bar{U}_t(y) = E\bar{Z}_t(y)/A(t), \\ V_t(y) &= 1 - \bar{V}_t(y), \quad \bar{V}_t(y) = E\bar{Y}_t(y)/N(t) \end{aligned} \tag{A.1}$$

Then the limiting residual life-time distributions are defined as

$$U(y) = 1 - \bar{U}(y) = 1 - \lim_{t \rightarrow \infty} \bar{U}_t(y), \quad V(y) = 1 - \bar{V}(y) = 1 - \lim_{t \rightarrow \infty} \bar{V}_t(y) \quad (A.2)$$

Note that  $E\bar{Z}_t(y) = A(t+y) - A(t+y, y)$  and  $E\bar{Y}_t(y) = N(t+y) - N(t+y, y)$ . Therefore

$$\begin{aligned} \bar{U}_t(y) &= [A(t+y) - A(t+y, y)]/A(t) = [A(t+y)/A(t)]\{1 - A(t+y, y)/A(t+y)\}, \\ \bar{V}_t(y) &= [N(t+y) - N(t+y, y)]/N(t) = [N(t+y)/N(t)]\{1 - N(t+y, y)/N(t+y)\}. \end{aligned} \quad (A.3)$$

Let us first consider the Markov case of Section 3.5. Then from (41) and (A.3) it follows

$$U_t(y) = 1 - \bar{U}_t(y) = 1 - e^{-\lambda y}, \quad y \geq 0. \quad (A.4)$$

Therefore, the Markov branching process (without immigration) has an exponential (with parameter  $\lambda$ ) *stationary* residual life-time distribution which is of the same form as the life-time distribution. On the other hand, from (44) and (A.3) we have

$$\begin{aligned} \bar{V}_t(y) &= e^{-m\lambda y} \{1 - e^{-\lambda(1-m)(t+y)}\} / \{1 - e^{-\lambda(1-m)t}\}, \quad m < 1; \\ \bar{V}_t(y) &= e^{-\lambda y} (1 + y/t), \quad m = 1; \\ \bar{V}_t(y) &= e^{-m\lambda y} \{e^{\lambda(m-1)y} - e^{-\lambda(m-1)t}\} / \{1 - e^{-\lambda(m-1)t}\}, \quad m > 1. \end{aligned}$$

Then the limiting residual life-time distribution is given by

$$V(y) = 1 - e^{-m\lambda y}, \quad m < 1 \quad (A.5)$$

$$V(y) = 1 - e^{-\lambda y}, \quad m \geq 1. \quad (A.6)$$

*Remark 1.* In the Markov branching process with homogeneous Poisson immigration, the residual life-time distributions depend of the critical parameter  $m$ . In the critical and supercritical cases, the limiting residual life-time distribution is the same as that in the process without immigration, while in the subcritical case it is exponential with parameter  $m\lambda$ .

Let us now consider the residual life-time distributions for the Bellman-Harris branching processes with and without immigration under conditions (29) and (30) of Section 3.4. Using formulas (31), (32) and (A.3), we arrive at the following limiting residual life-time distribution in this case:

$$U(y) = 1 - \frac{e^{\alpha y} \int_y^\infty e^{-\alpha u} (1 - G(u)) du}{\int_0^\infty e^{-\alpha u} (1 - G(u)) du}. \quad (A.7)$$

Note that in the critical case  $\alpha = 0$  the limiting distribution assumes the form

$$U(y) = \frac{\int_0^y (1 - G(u)) du}{\int_0^\infty (1 - G(u)) du}. \quad (A.8)$$

For the Bellman-Harris process with homogeneous Poisson immigration, proceeding from (A.3) and using formulas (32) – (35), we have

(i) If  $\alpha \geq 0$  then

$$V(y) = 1 - \frac{e^{\alpha y} \int_y^\infty e^{-\alpha u} (1 - G(u)) du}{\int_0^\infty e^{-\alpha u} (1 - G(u)) du}; \quad (A.9)$$

(i) If  $\alpha < 0$  then

$$V(y) = \frac{\int_0^y e^{-\alpha u} (1 - G(u)) du}{\int_0^\infty e^{-\alpha u} (1 - G(u)) du}. \quad (A.10)$$

*Remark 2.* Note that in the critical and supercritical cases the limiting residual life-time distributions for the Bellman-Harris process with and without immigration are identical. Setting  $G(t) = 1 - e^{-\lambda t}$ ,  $t \geq 0$ , in the limiting distributions (A.7) – (A.10), it is easy to obtain the corresponding stationary distribution given by (A.4) or the limiting distributions (A.5) and (A.6) in the Markov case.

## Legends to Figures

Figure 1. A typical structure of renewing cell systems with  $r_1, r_2, r_3, r_4, r_5$  representing the rates of transition between different cell types.

Figure 2. The evolution of glia-forming cells in the central nervous system. Notation:  $T_0$  – stem cell,  $T_1$  – GRP cell,  $T_2$  – O-2A/OPC,  $T_3$  – oligodendrocyte.

Figure 3. The evolution of leukemic cells. Notation:  $T_0$  – LSC,  $T_1$  – LP,  $T_2$  – LB.

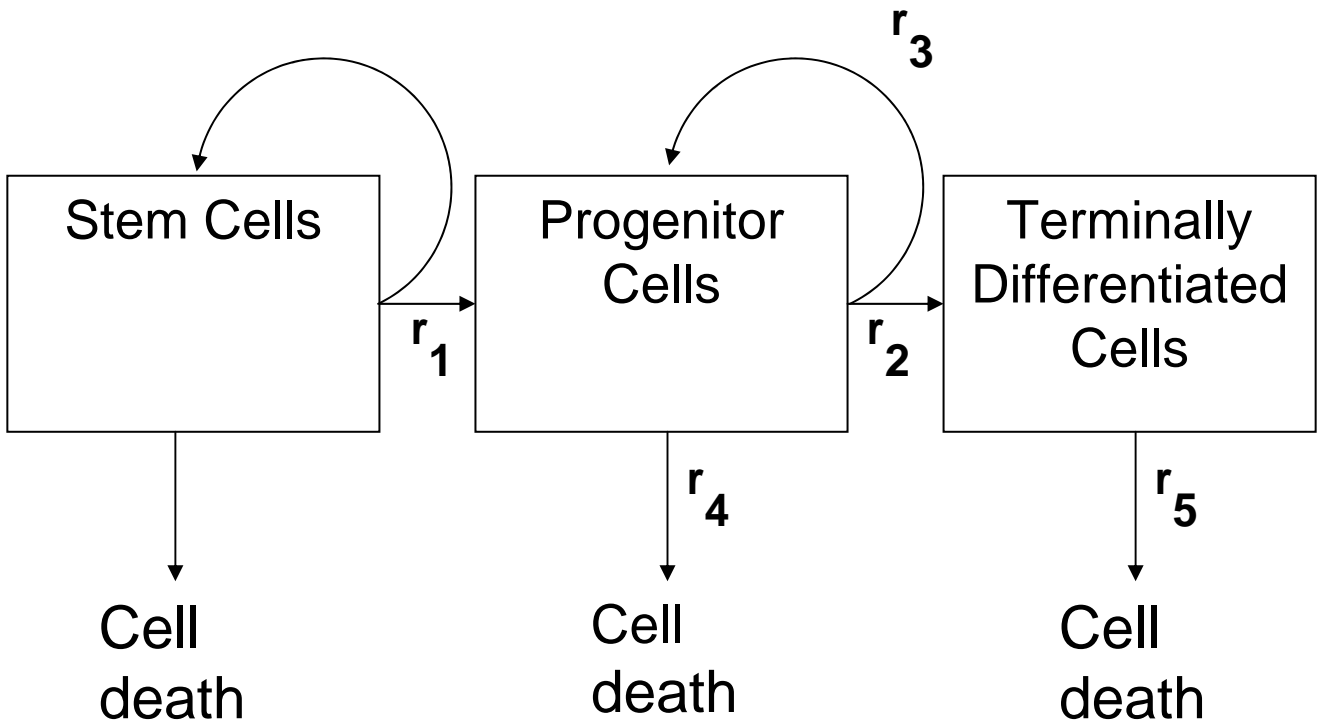


Figure 1

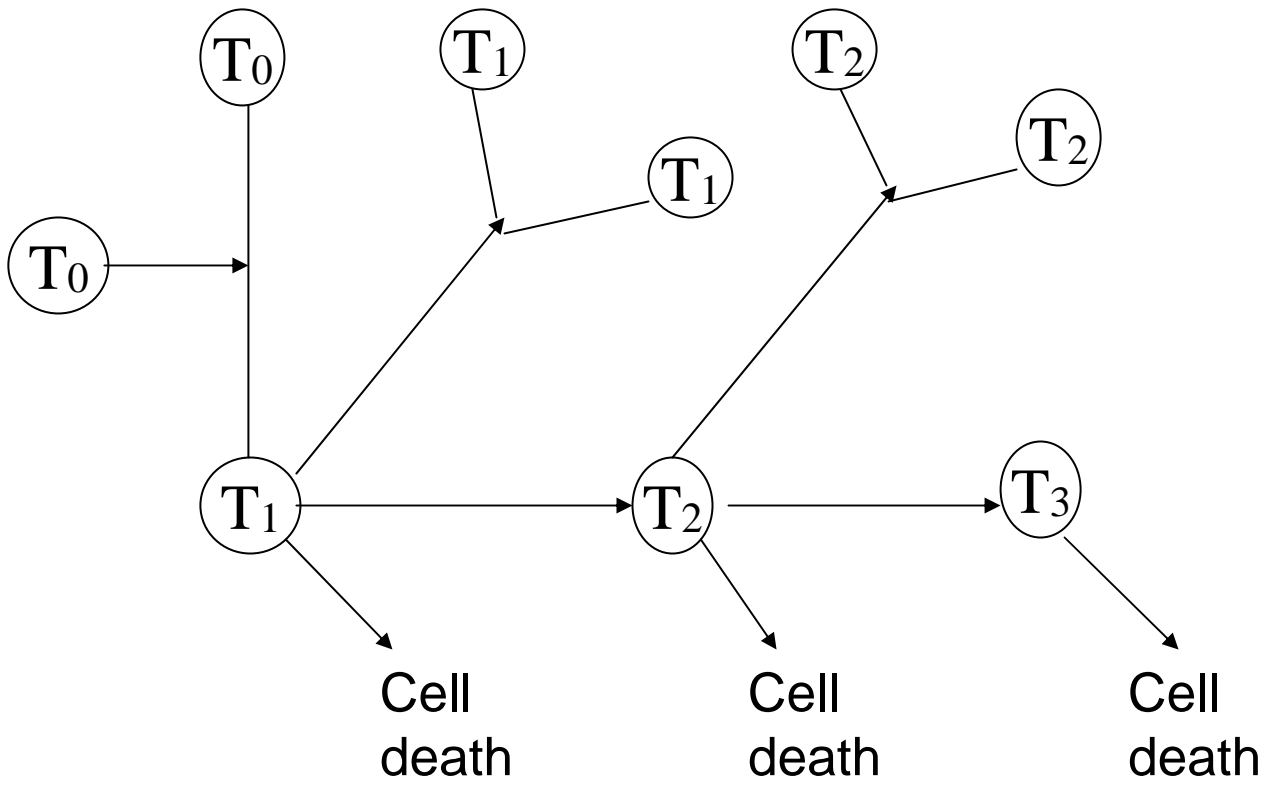


Figure 2

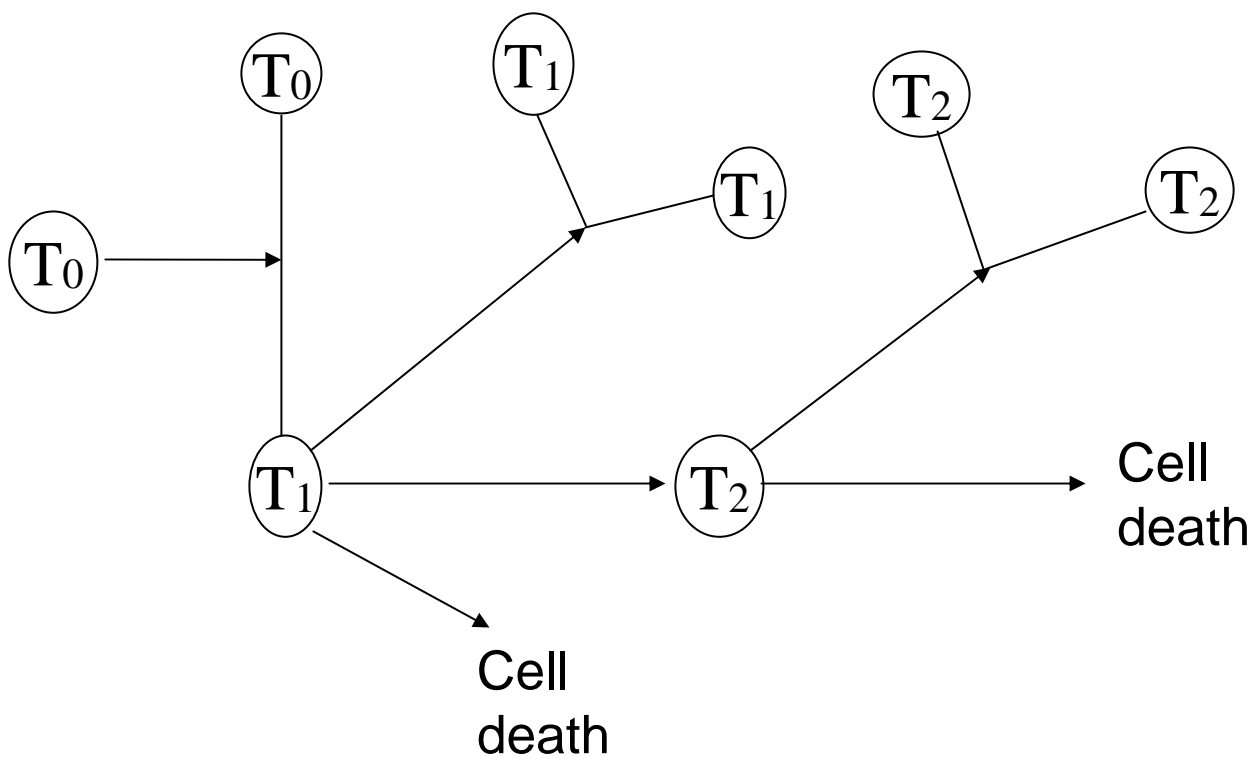


Figure 3