J. Appl. Prob. **52,** 864–876 (2015)
Printed in England

A MULTITYPE INFINITE-ALLELE BRANCHING PROCESS WITH APPLICATIONS TO CANCER EVOLUTION

THOMAS O. MCDONALD,* ** *Rice University*
MAREK KIMMEL,* *** *Rice University and Silesian University of Technology*

Abstract

We extend the infinite-allele simple branching process of Griffiths and Pakes (1988) allowing the offspring to change types and labels. The model is developed and limit theorems are given for the growth of the number of labels of a specific type. We also discuss the asymptotics of the frequency spectrum. Finally, we present an application of the model's use in tumorigenesis.

Keywords: Infinite-allele branching process; multitype BGW process; modeling cancer evolution; population genetics; tumorigenesis

2010 Mathematics Subject Classification: Primary 60J80
Secondary 60J85

1. Introduction

The aim of this paper is to introduce a multitype version of the infinite-allele Bienaymé–Galton–Watson (BGW) process introduced by Griffiths and Pakes [5]. We extend the process from individuals of a single type to a k-type process to allow differing growth parameters between types. We define $\{\boldsymbol{Z}(n), n = 0, 1, 2, \dots\}$ as the k-dimensional stochastic process where $\boldsymbol{Z}(n) = (Z_1(n), \dots, Z_k(n))$ is the number of particles of each type in the nth generation, $\mathcal{G}_n$. The process $\{\boldsymbol{Z}(n)\}$ is a typical multitype BGW process where each type represents a different set of the so-called driver mutations which may confer a growth advantage to the mutant cell clone [4]. Probabilities of driver mutations occurring are then represented in the offspring probability generating function (PGF) of $\{\boldsymbol{Z}(1)\}$, $\boldsymbol{f}(\boldsymbol{s}) = (f_1(\boldsymbol{s}), \dots, f_k(\boldsymbol{s}))$. That is, given a type i individual, the offspring PGF is

$$f_i(\boldsymbol{s}) = \sum_{j_1,\dots j_k \geq 0} p_i(j_1, \dots, j_k) s_1^{j_1} \dots s_k^{j_k}, \qquad s_i \in [0, 1],$$

where $p_i(j_1, \dots, j_k) = \mathbb{P}[\boldsymbol{Z}(1) = (j_1, \dots, j_k) \mid \boldsymbol{Z}(0) = \boldsymbol{e}_i]$. The transition probabilities of splitting into different types in the PGF describe driver mutation probabilities. We denote the mean offspring matrix by $\boldsymbol{M}$. Let ρ be the spectral radius of $\boldsymbol{M}$ with left and right eigenvectors $\boldsymbol{v}$ and $\boldsymbol{u}$. Results concerning differences in driver mutations, criticality, and asymptotics of $\boldsymbol{Z}(n)$ follow the standard theory of multitype processes and can be found in Athreya and Ney [1] and Mode [10].

Within this multitype process, we incorporate the possibility for newly-born offspring to have a passenger mutation with probability $\mu \in (0, 1)$ regardless of type. This extends the

Received 10 April 2014; revision received 8 October 2014.
* Postal address: Department of Statistics, Rice University, MS-138, P.O. Box 1892, Houston, TX 77251-1892, USA.
** Email address: thomas.o.mcdonald@rice.edu
*** Email address: kimmel@rice.edu

infinite-allele idea introduced by Griffiths and Pakes, where each individual branching process initiates an infinite-allele branching process. Since we can have a large number of passenger (selectively neutral) mutations that only affect heterogeneity, we do not distinguish between alleles that have a different set of passenger mutations in $\boldsymbol{Z}(n)$ or its PGF $\boldsymbol{f}^{(n)}(\boldsymbol{s})$. Instead we only track their count. To avoid confusion, we use the term *type* to distinguish individuals that differ with respect to driver mutations and have a particular type with respect to the branching process $\{\boldsymbol{Z}(n)\}$. That is $Z_i(n)$ and $Z_j(n)$ count different types within the population. When an individual has an offspring that undergoes a passenger mutation, we use the same terminology as Taïb [11] and say the offspring has a different *label*. Every passenger mutation event leads to a new and unique label. Thus, individuals can be distinguished by their *type* and *label*, but only the type influences growth rates.

We are interested in a particular quantity $\boldsymbol{K}(n) = (K_1(n), \ldots, K_k(n))$ the k-dimensional vector with $K_i(n)$ equal to the number of labels carried by i-type individuals in $\mathcal{G}_n$. The term $\{\boldsymbol{K}(n)\}$ is a stochastic process with $\boldsymbol{K}(0) = \boldsymbol{e}_i$ representing the only label present of the i-type ancestor counted by $\boldsymbol{Z}(0)$. We will also make use of the branching process $\{\tilde{\boldsymbol{Z}}(n)\}$ which we call the ancestor process of $\{\boldsymbol{Z}(n)\}$. The process $\tilde{\boldsymbol{Z}}(n)$ counts the number of individuals in generation n that have the same label as the ancestor, or never undergo a passenger mutation. We define the PGF for the ancestor offspring process of an i-type individual as

$$H_i(\boldsymbol{s}) \equiv \mathbb{E}[\boldsymbol{s}^{\tilde{\boldsymbol{Z}}(1)} \mid \tilde{\boldsymbol{Z}}(0) = \boldsymbol{e}_i] = f_i(\mu + (1-\mu)s_1, \ldots, \mu + (1-\mu)s_k).$$

The k-dimensional vector $\boldsymbol{H}(\boldsymbol{s})$ is the offspring PGF for the ancestor process. In both the normal individual process and the ancestor process the PGF in the nth generation is the nth iterate of the PGF, denoted by $\boldsymbol{f}^{(n)}(\boldsymbol{s})$ and $\boldsymbol{H}^{(n)}(\boldsymbol{s})$, respectively. We also denote the mean matrix of the ancestor process $\tilde{\boldsymbol{M}} = (1-\mu)\boldsymbol{M}$.

We count the number of labels for a particular type by counting individuals with specific characteristics. Define the indicator $I_{m,i,n} = 1$ if the mth i-type individual in $\mathcal{G}_n$ has a new label different from its parent. Also, define the indicator $J_{m,i,r,n-r}(j) = 1$ if some j-type individual in $\mathcal{G}_n$ has a label initiated by the mth i-type individual in $\mathcal{G}_r$. It follows then that

$$J_{m,i,0,n}(j) = \mathbf{1}_{\{\tilde{Z}_j(n)>0 \mid \tilde{\boldsymbol{Z}}(0)=\boldsymbol{e}_i\}}$$

and, furthermore,

$$\mathbb{E}[J_{m,i,0,n}(j)] = \mathbb{P}[\tilde{Z}_j(n) > 0 \mid \tilde{\boldsymbol{Z}}(0) = \boldsymbol{e}_i] = 1 - H_i^{(n)}(\mathbf{1} - \boldsymbol{e}_j),$$

where $\mathbf{1}$ denotes the vector of ones. If $\mathcal{F}_n$ is the natural filtration with respect to $\{\boldsymbol{Z}(n)\}$ then

$$\mathbb{E}[I_{m,i,r} J_{m,i,r,n-r}(j) \mid \mathcal{F}_r] = \mu(1 - H_i^{(n-r)}(\mathbf{1} - \boldsymbol{e}_j)).$$

We can express the number of labels of a certain type in terms of the number of individuals in the population that are ancestors to new labels and not yet extinct. Given a type α ancestor,

$$K_j(n) = J_{1,\alpha,0,n}(j) + \sum_{r=1}^{n} \sum_{i=1}^{k} \sum_{m=1}^{Z_i(r)} I_{m,i,n} J_{m,i,r,n-r}(j).$$

https://doi.org/10.1239/jap/1445543852 Published online by Cambridge University Press

The expectation given a type α ancestor is then

$$\mathbb{E}_\alpha[K_j(n)] = \mathbb{E}_\alpha[J_{1,\alpha,0,n}(j)] + \sum_{r=1}^{n}\sum_{i=1}^{k}\mathbb{E}_\alpha\left[\sum_{m=1}^{Z_i(r)} I_{m,i,n} J_{m,i,r,n-r}(j)\right]$$
$$= 1 - H_\alpha^{(n)}(\mathbf{1} - \boldsymbol{e}_j) + \mu\sum_{r=0}^{n-1}\boldsymbol{e}_\alpha^\top M^{n-r}(\mathbf{1} - \boldsymbol{H}^{(r)}(\mathbf{1} - \boldsymbol{e}_j)), \tag{1}$$

which we simplify using conditional expectation and rewriting the indices in the sum.

2. Irreducible $\boldsymbol{M}$

In this section we suppose the mean matrix $\boldsymbol{M}$ is irreducible. Also suppose ρ is the spectral radius of $\boldsymbol{M}$ with left and right eigenvectors $\boldsymbol{v}$ and $\boldsymbol{u}$ normalized so that $\boldsymbol{v}\boldsymbol{u} = 1$ and $\mathbf{1}^\top\boldsymbol{u} = 1$. The eigenvector $\boldsymbol{v}$ is a row vector, and $\boldsymbol{u}$ is a column vector. Let us define the constant

$$A_j = \mu\sum_{r=0}^{\infty}\boldsymbol{v}\rho^{-r}[\mathbf{1} - \boldsymbol{H}^{(r)}(\mathbf{1} - \boldsymbol{e}_j)].$$

Note that this constant is finite regardless of the criticality of ρ. This yields the following lemma about the limit of the expectation as given in (1).

Lemma 1. *Given an irreducible process starts with an ancestor of type α,*

$$\lim_{n\to\infty}\frac{\mathbb{E}_\alpha[K_j(n)]}{\mathbb{E}_\alpha[\boldsymbol{Z}(n)\boldsymbol{u}]} = A_j$$

Proof. First note that $\mathbb{E}_\alpha[\boldsymbol{Z}(n)\boldsymbol{u}] = \boldsymbol{e}_\alpha^\top\boldsymbol{M}^n\boldsymbol{u} = \boldsymbol{e}_\alpha^\top\rho^n\boldsymbol{u}$.

If $\rho > 1$ then $(1 - H_\alpha^{(n)}(\mathbf{1} - \boldsymbol{e}_j))\rho^{-n} \to 0$. If $\rho \le 1$ then $(1 - H_\alpha^{(n)}(\mathbf{1} - \boldsymbol{e}_j)) \to 0$ since extinction occurs almost surely (a.s.).

The second term of the sum in(1) can be rewritten as

$$\frac{\mu}{\boldsymbol{e}_\alpha^\top\rho^n\boldsymbol{u}}\sum_{r=0}^{n-1}\boldsymbol{e}_\alpha^\top\boldsymbol{M}^{n-r}(\mathbf{1} - \boldsymbol{H}^{(r)}(\mathbf{1} - \boldsymbol{e}_j)) = \frac{\mu\boldsymbol{e}_\alpha^\top\boldsymbol{M}^n}{\boldsymbol{e}_\alpha^\top\rho^n\boldsymbol{u}}\sum_{r=0}^{n-1}\boldsymbol{M}^{-r}(\mathbf{1} - \boldsymbol{H}^{(r)}(\mathbf{1} - \boldsymbol{e}_j)).$$

For irreducible $\boldsymbol{M}$, $\lim_{n\to\infty}(\boldsymbol{M}\rho^{-1})^n = \boldsymbol{u}\boldsymbol{v}$. Also,

$$1 - H_\alpha^{(n)}(\mathbf{1} - \boldsymbol{e}_j) \le \mathbb{E}_\alpha[\tilde{Z}_j(n)] = \boldsymbol{e}_\alpha^\top\tilde{\boldsymbol{M}}^n\boldsymbol{e}_j$$

by Markov's inequality. Since $\tilde{\boldsymbol{M}} = (1 - \mu)\boldsymbol{M}$, then $\boldsymbol{M}^{-r}\tilde{\boldsymbol{M}}^r = (1 - \mu)^r I$, so

$$\sum_{r=0}^{n-1}\boldsymbol{M}^{-r}(\mathbf{1} - \boldsymbol{H}^{(r)}(\mathbf{1} - \boldsymbol{e}_j)) \le \sum_{r=0}^{n-1}\boldsymbol{M}^{-r}\tilde{\boldsymbol{M}}^r\boldsymbol{e}_j$$
$$= \sum_{r=0}^{n-1}(1 - \mu)^r\boldsymbol{e}_j$$
$$\to \frac{1}{\mu}\boldsymbol{e}_j \quad \text{as } n \to \infty.$$

https://doi.org/10.1239/jap/1445543852 Published online by Cambridge University Press

This series converges absolutely, leading to

$$\lim_{n\to\infty}\frac{\mathbb{E}_\alpha[K_j(n)]}{\boldsymbol{e}_\alpha^\top\rho^n\boldsymbol{u}}=\frac{\mu\boldsymbol{e}_\alpha^\top}{\boldsymbol{e}_\alpha^\top\boldsymbol{u}}\boldsymbol{u}\boldsymbol{v}\sum_{r=0}^{\infty}\boldsymbol{M}^{-r}(\mathbf{1}-\boldsymbol{H}^{(r)}(\mathbf{1}-\boldsymbol{e}_j))$$
$$=\mu\sum_{r=0}^{\infty}\boldsymbol{v}\boldsymbol{M}^{-r}(\mathbf{1}-\boldsymbol{H}^{(r)}(\mathbf{1}-\boldsymbol{e}_j))$$
$$=\mu\sum_{r=0}^{\infty}\boldsymbol{v}\rho^{-r}(\mathbf{1}-\boldsymbol{H}^{(r)}(\mathbf{1}-\boldsymbol{e}_j))$$

giving the desired result for all $\rho>0$.

If we define $\Omega_a=\{\boldsymbol{Z}(n)>\mathbf{0},\ n=0,1,2,\dots\}$ as the set of nonextinction, then we can show the limiting behavior of $K_j(n)$ for supercritical processes ($\rho>1$) converges a.s. to A_j conditionally on nonextinction.

Theorem 1. *If $\mathbb{E}_i[Z_j(1)\log Z_j(1)]<\infty$ for all $1\le i\le k$, $1\le j\le k$, and $1<\rho<\infty$, and if the process is started by a α-type ancestor, then*

$$\lim_{n\to\infty}\frac{K_j(n)}{\boldsymbol{Z}(n)\boldsymbol{u}}=A_j\quad a.s.\ on\ \Omega_a.$$

Proof. Define the variable

$$K_j(r,n-r)=\sum_{i=1}^{k}\sum_{m=1}^{Z_i(r)}I_{m,i,r}J_{m,i,r,n-r}(j),$$

which is the number of new j-type labels in $\mathcal{G}_r$ that are still represented in $\mathcal{G}_n$. Note that $K_j(n)=\sum_{r=0}^{n}K_j(r,n-r)$. Then for any fixed n',

$$(\boldsymbol{Z}(n)\boldsymbol{u})^{-1}\sum_{r=1}^{n'-1}K_j(r,n-r)\le(\boldsymbol{Z}(n)\boldsymbol{u})^{-1}\sum_{r=1}^{n'-1}Z_j(r)\longrightarrow 0\quad\text{a.s. on }\Omega_a.$$

We can represent $K_j(n)$ as a sum and show that each of the summands converge. First, we show that

$$\frac{\sum_{r=n'}^{n}K_j(r,n-r)}{\boldsymbol{Z}(n)\boldsymbol{u}}=\frac{\sum_{r=0}^{n-n'}K_j(n-r,r)}{\boldsymbol{Z}(n)\boldsymbol{u}}=\frac{\sum_{r=0}^{n-n'}K_j(n-r,r)}{\boldsymbol{Z}(n-r)\boldsymbol{u}}\left(\frac{\boldsymbol{Z}(n-r)\boldsymbol{u}}{\boldsymbol{Z}(n)\boldsymbol{u}}\right).\tag{2}$$

Now, conditioning on $\boldsymbol{Z}(n-r)$, $\mathbb{E}[K_j(n-r,r)\mid\boldsymbol{Z}(n-r)]=\mu(\boldsymbol{Z}(n-r)[\mathbf{1}-\boldsymbol{H}^{(n-r)}(\mathbf{1}-\boldsymbol{e}_j)])$ is a sum of independent and identically distributed random variables. Noting that $\boldsymbol{Z}(n)/(\boldsymbol{Z}(n)\boldsymbol{u})\to\boldsymbol{v}$ a.s. on Ω_a as given in Athreya and Ney [1, Theorems 1 and 4, p. 193] and proved by Kurtz *et al.* [9], we can use a strong law for random sums to obtain almost sure convergence $K_l(n-r,r)/(\boldsymbol{Z}(n)\boldsymbol{u})\to\mu\boldsymbol{v}[\mathbf{1}-\boldsymbol{H}^{(n-r)}(\mathbf{1}-\boldsymbol{e}_l)]$ on Ω_a as $n\to\infty$ and $r=O(1)$.

Hoppe [7, Theorem 2.1] showed that for $1<\rho<\infty$, there exists a sequence of positive vectors, $\{\boldsymbol{c_n}\}$ and scalars $\{\gamma_n\}=\{\boldsymbol{v}\boldsymbol{c_n}\}$ such that for each α, if $\boldsymbol{Z}(\mathbf{0})=\boldsymbol{e}_\alpha$ then

$$\lim_{n\to\infty}\boldsymbol{Z}(n)\boldsymbol{c_n}=W_\alpha\quad\text{a.s.},\tag{3}$$
$$\frac{\lim_{n\to\infty}\gamma_n}{\gamma_{n+1}}=\rho,\tag{4}$$

https://doi.org/10.1239/jap/1445543852 Published online by Cambridge University Press

$$\frac{\lim_{n\to\infty} \boldsymbol{c_n}}{\gamma_n} = \boldsymbol{u}, \tag{5}$$

$$\lim_{n\to\infty} \gamma_n \boldsymbol{Z}(n) = W_\alpha \boldsymbol{v} \quad \text{a.s.}, \tag{6}$$

where $\boldsymbol{W} = (W_1, \ldots, W_K)$ is a nonnegative random variable if $\mathbb{E}[Z_i(1) \log Z_i(1)] < \infty$ for $i = 1, \ldots, k$. Now, since $\gamma_n \boldsymbol{Z}(n)\boldsymbol{u} \to W_\alpha$ a.s. by (6), we make use of (3)–(5) to obtain

$$\frac{\boldsymbol{Z}(n-1)\boldsymbol{u}}{\boldsymbol{Z}(n)\boldsymbol{u}} \sim \frac{W_\alpha \gamma_n}{W_\alpha \gamma_{n-1}} \longrightarrow \rho^{-1},$$

implying

$$\frac{\boldsymbol{Z}(n-r)\boldsymbol{u}}{\boldsymbol{Z}(n)\boldsymbol{u}} \longrightarrow \rho^{-r} \quad \text{a.s. as } n \to \infty.$$

Since

$$\frac{K_j(n-r, r)}{\boldsymbol{Z}(n-r)\boldsymbol{u}} \le \frac{Z(n-r)}{\boldsymbol{Z}(n-r)\boldsymbol{u}},$$

each summand of the right-hand side of (2) is dominated by ρ^{-r}. Thus, the series is dominated by the geometric series that converges to $(1-\rho)^{-1}$, so we can use the dominated convergence theorem to show the series on the right-hand side of (2) converges to A_j a.s. on Ω_a as $n \to \infty$. The assertion follows.

3. Reducible M

We now remove the assumption of irreducibility from the process to model more realistic scenarios. Mutations usually have low probability of being reversed, so if we assume the probability is 0 (as often is the case), we want to understand how the number of labels grows within each type. We are able to group the types of the branching process into equivalence classes that form irreducible subprocesses. This allows a reordering of matrix $\boldsymbol{M}$ as a block lower triangular matrix with blocks along the diagonal being irreducible. Results about the process can then be ascertained based on results for the blocks.

Define the equivalence classes $\{C_a\}_{a=1,\ldots,l}$ as

$$C_a = \{i, j \in 1, 2, \ldots, k; m_{i,j}^{(n_1)} > 0 \text{ and } m_{j,i}^{(n_2)} > 0 \text{ for some } n_1 \text{ and } n_2\}.$$

That is, the equivalence classes are created by separating types into groups where each type in the group communicates with every other type in the same group [8]. We are able to order the indices of types and permute the mean matrix $\boldsymbol{M}$ according to the equivalence classes by imposing an order on the classes such that if $b > a$, then $m_{i,j}^{(n)} > 0$ for all $i \in C_b$ and $j \in C_a$. The resulting mean matrix after permutation is

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{M}_{1,1} & 0 & 0 & \ldots & 0 \\ \boldsymbol{M}_{2,1} & \boldsymbol{M}_{2,2} & 0 & \ldots & 0 \\ \boldsymbol{M}_{3,1} & \boldsymbol{M}_{3,2} & \boldsymbol{M}_{3,3} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{M}_{l,1} & \boldsymbol{M}_{l,2} & \boldsymbol{M}_{l,3} & \ldots & \boldsymbol{M}_{l,l} \end{bmatrix}$$

with $\boldsymbol{M}_{a,b} = (m_{i,j}), i \in C_a, j \in C_b$. If we limit a process to types within a subclass, then $\boldsymbol{Z}_{C_a}(n) = \{Z_i(n), i \in C_a\}$ is an irreducible subprocess of $\boldsymbol{Z}(n)$ having mean offspring

https://doi.org/10.1239/jap/1445543852 Published online by Cambridge University Press

matrix $\boldsymbol{M}_{a,a}$. Given an ancestor with type $\alpha \in C_a$, results about $\boldsymbol{Z}_{C_a}(n)$ and $\boldsymbol{K}_{C_a}(n)$ follow those of the previous section. We show here the results for a 2-class example and note that examples with more than 2 classes can be developed analogously.

Suppose we have a reducible BGW process $\boldsymbol{Z}(n)$ with mean matrix

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{M}_{1,1} & 0 \\ \boldsymbol{M}_{2,1} & \boldsymbol{M}_{2,2} \end{bmatrix}$$

with irreducible $\boldsymbol{M}_{1,1}$ and $\boldsymbol{M}_{2,2}$, and $\boldsymbol{M}_{2,1} \neq \boldsymbol{0}$. Let $\boldsymbol{M}_{i,i}$ have spectral radius ρ_i with associated left and right eigenvectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ for $i = 1, 2$. Let ρ be the spectral radius of $\boldsymbol{M}$, which is the maximum of ρ_1 and ρ_2. If $\rho = \rho_1 > \rho_2$ then the eigenvectors associated with ρ are $\boldsymbol{v} = (\boldsymbol{v}_1, 0)$ and $\boldsymbol{u} = (\boldsymbol{u}_1, (\rho I - \boldsymbol{M}_{2,2})^{-1}\boldsymbol{M}_{2,1}\boldsymbol{u}_1)$. If $\rho = \rho_2 > \rho_1$ then $\boldsymbol{v} = (\boldsymbol{v}_2\boldsymbol{M}_{2,1}(\rho I - \boldsymbol{M}_{1,1})^{-1}, \boldsymbol{v}_2)$ and $\boldsymbol{u} = (0, \boldsymbol{u}_2)$. These are used as eigenvectors when we show the limits hold for the reducible cases. We present an analogous lemma to Lemma 1 for the expectation of the number of labels $K_j(n)$ in a 2-class process with the mean matrix as above. We also limit ourselves to an ancestor of a type in the class C_2 to avoid trivial results that can arise. Also, we expect ancestors of cell populations to have no somatic mutations, but have the ability to gain mutations. We do not expect mutant cells to give rise to daughter cells without those mutations since the reversing of mutations is very rare. Because of this, we require the reducibility assumption for $\boldsymbol{M}$. Another case can occur when $\rho_1 = \rho_2$. Different convergence results exist in this situation which can be determined analogously based on Kesten and Stigum [8, Theorem 2.3]. This situation is less likely to occur in cancer evolution, so it is not discussed here.

Lemma 2. *Suppose that ρ is the spectral radius of $\boldsymbol{M}$ and it is simple. Given a reducible BGW process with an ancestor of type $\alpha \in C_2$,*

$$\lim_{n\to\infty} \frac{\mathbb{E}_\alpha[K_j(n)]}{\mathbb{E}_\alpha[\boldsymbol{Z}(n)\boldsymbol{u}]} = A_j$$

with $\boldsymbol{u}$ and $\boldsymbol{v}$ defined as above based on whether $\rho = \rho_1$ or $\rho = \rho_2$, and

$$A_j = \mu \sum_{r=0}^{\infty} \boldsymbol{v}\rho^{-r}[\boldsymbol{1} - \boldsymbol{H}^{(r)}(\boldsymbol{1} - \boldsymbol{e}_j].$$

Proof. Under the above conditions, $\lim(\boldsymbol{M}/\rho)^n = \boldsymbol{u}\boldsymbol{v}$ as $n \to \infty$ where $\boldsymbol{v}\boldsymbol{u} = 1$ [3]. The remaining calculations of Lemma 1 then hold with $\boldsymbol{u}$ and $\boldsymbol{v}$ determined by whether $\rho = \rho_1$ or $\rho = \rho_2$.

The almost sure convergence of $K_j(n)$ is also extended to the reducible case. Suppose again we consider a α-type ancestor with $\alpha \in C_2$. Let Ω_a still be defined as the set of nonextinction.

Theorem 2. *Assume that $\mathbb{E}[Z_i(1)\log Z_i(1) \mid \boldsymbol{Z}(0) = e_\alpha] \leq \infty$ for $i = 1, 2, \ldots, k$. Also, let ρ be the spectral radius of $\boldsymbol{M}$ and assume it is simple. Then*

$$\lim_{n\to\infty} \frac{K_j(n)}{\boldsymbol{Z}(n)\boldsymbol{u}} = A_j \quad \text{a.s. on } \Omega_a$$

with $\boldsymbol{u}$ and $\boldsymbol{v}$ defined as above.

We first introduce a corollary to Kesten and Stigum's almost sure convergence of $\boldsymbol{Z}(n)$ that we use in our proof.

https://doi.org/10.1239/jap/1445543852 Published online by Cambridge University Press

Corollary 1. *Suppose ρ is the spectral radius of $\boldsymbol{M}$, the mean matrix for $\boldsymbol{Z}(n)$ with eigenvalues $\boldsymbol{u}$ and $\boldsymbol{v}$ as given above depending on whether $\rho = \rho_1$ or $\rho = \rho_2$. Then*

$$\lim_{n\to\infty} \frac{\boldsymbol{Z}(n)}{\boldsymbol{Z}(n)\boldsymbol{u}} = \boldsymbol{v} \quad \text{a.s. on } \Omega_a.$$

Proof. Theorem 2.1 of Kesten and Stigum [8] states that

$$\lim_{n\to\infty} \frac{\boldsymbol{Z}(n)}{\rho^n} = w \cdot \boldsymbol{v} \quad \text{a.s.},$$

where w is a scalar random variable. This implies that

$$\lim_{n\to\infty} \frac{\rho^n}{\boldsymbol{Z}(n)\boldsymbol{u}} = w^{-1}$$

and, furthermore,

$$\lim_{n\to\infty} \frac{\boldsymbol{Z}(n)}{\boldsymbol{Z}(n)\boldsymbol{u}} = \lim_{n\to\infty} \frac{\rho^n \boldsymbol{Z}(n)}{\rho^n \boldsymbol{Z}(n)\boldsymbol{u}} = \boldsymbol{v} \quad \text{a.s. on } \Omega_a.$$

Proof of Theorem 2. Since $\alpha \in C_2$, the proof is similar to that of Theorem 1 with $\boldsymbol{u}$ and $\boldsymbol{v}$ defined as above based on whether $\rho = \rho_1 > \rho_2$ or $\rho = \rho_2 > \rho_1$. We use the reducible versions of the Kesten and Stigum theorem [8, Theorem 1.1] in place of the irreducible case to show almost sure convergence of the number of individuals. The calculations from the proof of Theorem 1 above then hold with the modified $\boldsymbol{u}$ and $\boldsymbol{v}$.

4. Frequency spectrum

Let $\alpha_i(j, n)$ be the number of i-type labels in generation n represented by j individuals currently living in generation n. We will denote the expectation of this term as the frequency spectrum for type i. The term $\alpha_i(j, n)$ can be expressed as a sum of indicators via an approach similar to determining the total number of labels. Define $I_{j,i}(n)$ as the indicator that the ancestor has j i-type descendants with the same label in generation n and $I_{j,i,l,k}(r, n-r)$ as the indicator that the lth k-type individual in generation r is a new label and has j i-type descendants with the same label in generation n. Given an ancestor of type α, $\mathbb{E}[I_{j,i}(n)] = \mathbb{P}[\tilde{Z}_i(n) = j \mid \tilde{Z}_\alpha(0) = 1]$, which we denote by $q^{(n)}_{\alpha,i}(j)$. The values of $q^{(n)}_{\alpha,i}(j)$ over i and j make up the coefficients to $H^{(n)}_\alpha(\boldsymbol{s})$. We can write $\alpha_i(j, n)$ in terms of the previous indicators

$$\alpha_{j,i}(n) = I_{j,i}(n) + \sum_{r=1}^{n} \sum_{k=1}^{K} \sum_{l=1}^{Z_k(r)} I_{j,i,l,k}(r, n-r)$$

allowing us to derive an expression for the frequency spectrum, $\phi_i(j, n) \equiv \mathbb{E}[\alpha_i(j, n)]$. The frequency spectrum can be simplified to

$$\begin{aligned}\phi_i(j, n) &= \mathbb{E}\Big[I_{j,i}(n) + \sum_{r=1}^{n} \sum_{k=1}^{K} \sum_{l=1}^{Z_k(r)} I_{j,i,l,k}(r, n-r)\Big] \\ &= q^{(n)}_{\alpha,i}(j) + \sum_{r=1}^{n} \sum_{k=1}^{K} \mu q^{(n-r)}_{k,i}(j) \boldsymbol{e}'_\alpha \boldsymbol{M}^r \boldsymbol{e}_k\end{aligned}$$

https://doi.org/10.1239/jap/1445543852 Published online by Cambridge University Press

using conditional expectation. Now denote the vector $\boldsymbol{q}_i^{(r)}(j) = [q_{1,i}^{(r)}(j), \ldots, q_{K,i}^{(r)}(j)]$. Then

$$\begin{aligned}\phi_i(j,n) &\equiv \mathbb{E}[\alpha_i(j,n)] \\ &= q_{\alpha,i}^{(n)}(j) + \sum_{r=1}^{n} \mu \boldsymbol{e}_\alpha' \boldsymbol{M}^r \boldsymbol{q}_i^{(n-r)}(j) \\ &= q_{\alpha,i}^{(n)}(j) + \mu \sum_{r=0}^{n-1} \boldsymbol{e}_\alpha' \boldsymbol{M}^{n-r} \boldsymbol{q}_i^{(r)}(j). \end{aligned} \tag{7}$$

Theorem 3. *Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be the right and left eigenvectors associated with ρ. Then*

$$\lim_{n\to\infty} \frac{\phi_i(j,n)}{\boldsymbol{Z}(n)\boldsymbol{u}} = \sum_{r=0}^{\infty} \mu \rho^{-r} \boldsymbol{v} \boldsymbol{q}_i^{(r)}(j), \tag{8}$$

$$\Psi(i,j) = \lim_{n\to\infty} \frac{\phi_i(j,n)}{\mathbb{E}[K_i(n)]} = \frac{\sum_{r=0}^{\infty} \mu \rho^{-r} \boldsymbol{v} \boldsymbol{q}_i^{(r)}(j)}{A_i}.$$

Proof. The proof is essentially the same as that of Lemma 1. We use the fact that $q_{\alpha,i}^{(n)}(j) \to 0$ as $n \to \infty$ in (7), and the sum is bounded.

In this case, $\Psi(i,j)$ is the long-run frequency of i-type labels having j individuals. This provides an idea of the distribution of labels having different individuals.

5. Proof of concept simulations

We consider two different 4-type branching processes with similar PGFs to illustrate the almost sure convergence results. Each process contains cells undergoing reproduction via binary fission or death, and the probability of a new allele at each generation is $\mu = 5 \times 10^{-4}$. The first process is irreducible with PGF

$$\begin{aligned} f_1(\boldsymbol{s}) &= 0.45 + 0.03 s_1 s_2 + 0.02 s_1 s_3 + 0.50 s_1^2, \\ f_2(\boldsymbol{s}) &= 0.51 + 0.06 s_1 s_2 + 0.04 s_2 s_3 + 0.39 s_2^2, \\ f_3(\boldsymbol{s}) &= 0.56 + 0.04 s_2 s_3 + 0.05 s_3 s_4 + 0.35 s_3^2, \\ f_4(\boldsymbol{s}) &= 0.50 + 0.03 s_2 s_4 + 0.06 s_3 s_4 + 0.40 s_4^2. \end{aligned}$$

The mean matrix is

$$\boldsymbol{M} = \begin{bmatrix} 1.05 & 0.03 & 0.02 & 0 \\ 0.06 & 0.88 & 0.04 & 0 \\ 0 & 0.04 & 0.79 & 0.05 \\ 0 & 0.03 & 0.07 & 0.90 \end{bmatrix}.$$

The spectral radius of this process is $\rho = 1.0617$ with left and right eigenvectors $\boldsymbol{v} = [1.3974, 0.2728, 0.1554, 0.0480]$ and $\boldsymbol{u} = [0.6637, 0.2291, 0.0451, 0.0620]^\top$. Thus, the process is supercritical, and growth is expected. Numerically evaluating $\boldsymbol{A}$, we obtain

$$\boldsymbol{A} = [0.0035, 0.0016, 0.0012, 0.0005]^\top.$$

We ran 100 simulations of the process beginning with a single type 1 ancestor. The results for the sample paths for each type is shown in Figure 1. We condition on nonextinction by removing

https://doi.org/10.1239/jap/1445543852 Published online by Cambridge University Press

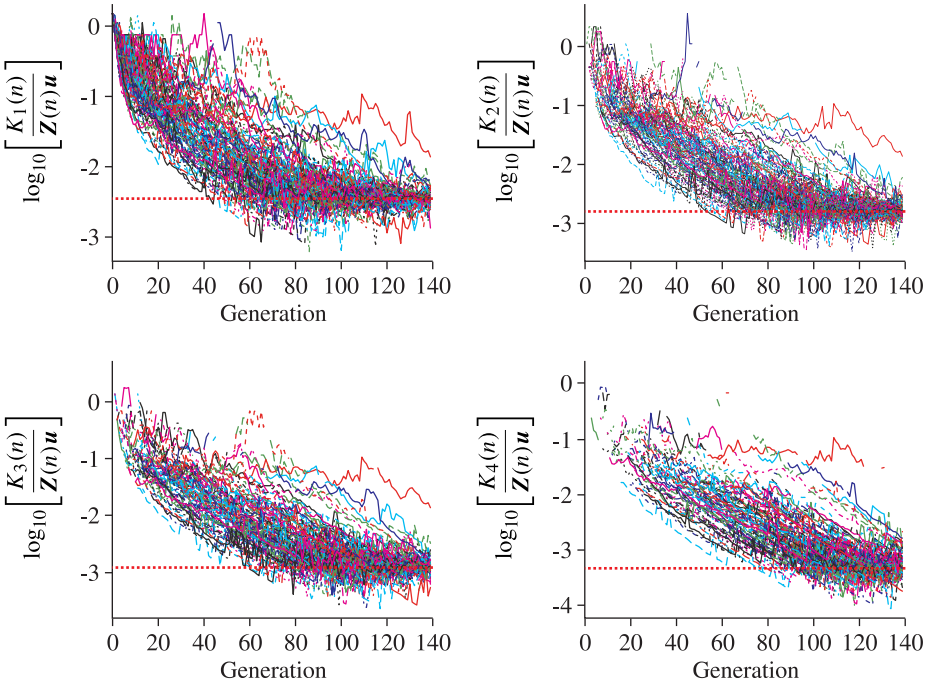

FIGURE 1: The paths of the process $\boldsymbol{K}(\boldsymbol{n})$ for 100 simulations shows convergence to $A_i, i = 1, 2, 3, 4$, which is given by the horizontal dashed line. We scale the paths with a $\log_{10}$ transformation to see convergence over the 140 generations.

simulations that become extinct and only using the remaining simulations. A horizontal dashed line is superimposed at the respective value of A_i according to the type, i. Over 140 generations, we see convergence of nearly all paths to the limit, A_i.

We created a similar 4-type process, but adjusted the probabilities so that types 3 and 4 form a class that can feed into types 1 and 2, but not in the other direction. The PGF for the process is

$$\begin{aligned} f_1(\boldsymbol{s}) &= 0.47 + 0.03s_1s_2 + 0.5s_1^2, \\ f_2(\boldsymbol{s}) &= 0.51 + 0.1s_1s_2 + 0.39s_2^2, \\ f_3(\boldsymbol{s}) &= 0.54 + 0.02s_2s_3 + 0.09s_3s_4 + 0.35s_3^2, \\ f_4(\boldsymbol{s}) &= 0.4 + 0.08s_2s_4 + 0.07s_3s_4 + 0.45s_4^2. \end{aligned}$$

The mean matrix is

$$\boldsymbol{M} = \begin{bmatrix} 1.03 & 0.03 & 0 & 0 \\ 0.1 & 0.88 & 0 & 0 \\ 0 & 0.02 & 0.81 & 0.09 \\ 0 & 0.08 & 0.07 & 1.05 \end{bmatrix}.$$

After breaking the matrices up into submatrices and determining the spectral radii of each, we find that the process is supercritical with spectral radius $\rho = 1.0739$ since $\rho_2 > \rho_1$.

https://doi.org/10.1239/jap/1445543852 Published online by Cambridge University Press

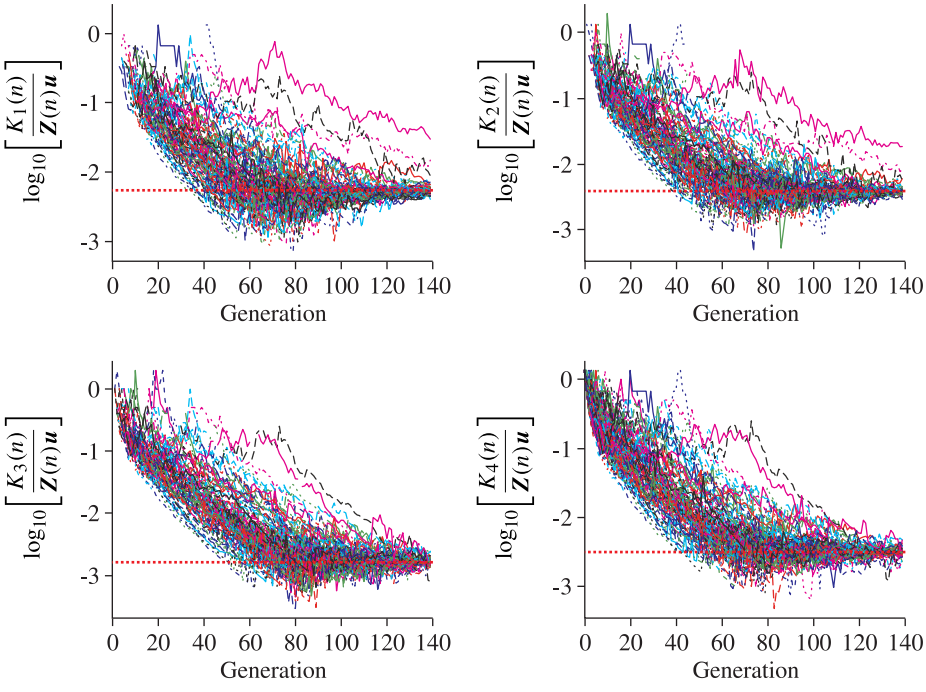

FIGURE 2: The paths of the process $\boldsymbol{K}(\boldsymbol{n})$ for 100 simulations in a reducible BGW shows convergence to $A_i, i = 1, 2, 3, 4$ which is given by the dashed line. We scale the paths with a $\log_{10}$ transformation to see convergence over the 140 generations.

The results of Kesten and Stigum lead to vectors $\boldsymbol{v} = [1.9053, 0.8359, 0.3262, 1.2298]$ and $\boldsymbol{u} = [0, 0, 0.2543, 0.7457]^\top$. We numerically evaluate $\boldsymbol{A}$ to obtain

$$\boldsymbol{A} = [0.0055, 0.0039, 0.0017, 0.0032]^\top.$$

The results of 100 simulations are shown in Figure 2 with a dashed horizontal line at the value of A_i. Note that these simulations are performed under the condition of nonextinction and are initiated with a single type 4 ancestor in generation 0. In Figure 2, we show similar results in the reducible case that holds by adjusting the eigenvectors according to Kesten and Stigum's results.

Finally, we show the results of the convergence of the frequency spectrum, Theorem 3, in a 2-type simulation starting with a single type 1 ancestor. Again, we set the probability of a new label to $\mu = 5 \times 10^{-4}$. Our PGF for this process is

$$f_1(\boldsymbol{s}) = 0.5s_1 + 0.2s_1s_2 + 0.3s_1^2, \qquad f_2(\boldsymbol{s}) = 0.5s_2 + 0.4s_1s_2 + 0.1s_2^2.$$

The process is supercritical with mean matrix

$$\boldsymbol{M} = \begin{bmatrix} 1.3 & 0.2 \\ 0.4 & 1.1 \end{bmatrix}$$

https://doi.org/10.1239/jap/1445543852 Published online by Cambridge University Press

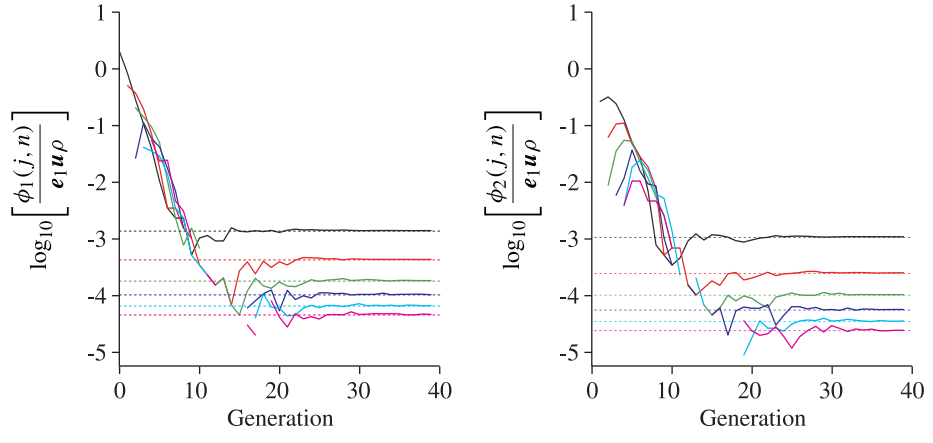

FIGURE 3: The simulation of the frequency spectrum for type-1 (*left*) and type-2 (*right*) individuals shows convergence to the analytical formula given by the horizontal lines for each number of individuals (ascending vertically each solid/dashed line represents 6–1 individuals).

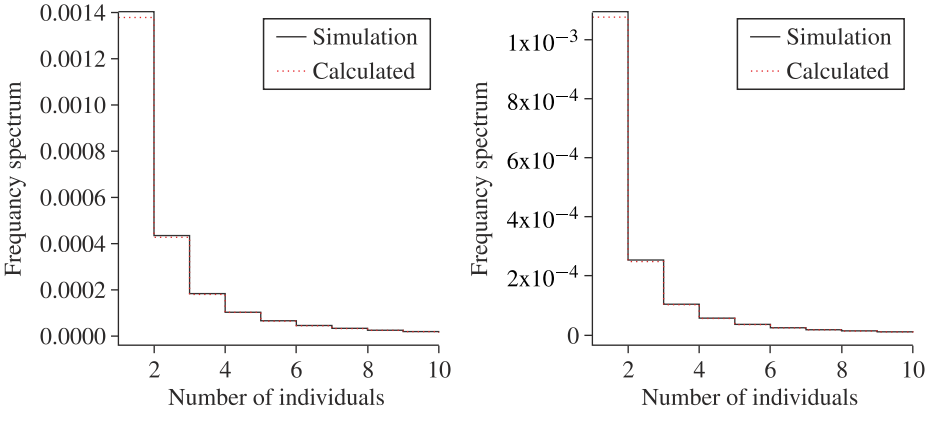

FIGURE 4: The results of the frequency spectrum from simulations after 40 generations is given for type-1 (*left*) and type-2 (*right*) individuals (solid line) along with the calculated curve (dashed line).

having spectral radius of $\rho = 1.5$ and eigenvectors $\boldsymbol{v} = [\frac{4}{3}, \frac{2}{3}]$ and $\boldsymbol{u} = [\frac{1}{2}, \frac{1}{2}]^\top$. The results for the convergence of the normalized frequency spectrum are shown in Figure 3. We scaled the process using a $\log_{10}$ transformation to better illustrate the convergence and difference in each curve. Each type of line represents the average number of labels represented by j individuals after 100 simulations. In each case, the value converges to the numerical solution given on the right-hand side of the theorem. The two plots refer to both types in the process. In Figure 4 we show the results in the 40th generation, after convergence has occurred. The results of simulations of the process represented by the left-hand side of (8) in Theorem 3 are shown as a solid line and the results of calculating the right-hand side to show convergence are shown as a dashed line. In both cases, there is very little error after the 40 generations.

https://doi.org/10.1239/jap/1445543852 Published online by Cambridge University Press

6. Applications to cancer evolution

The current view of cancer progression is that the multistep accumulation of somatic mutations leads to the transformation of healthy cells into malignant cancer cells with higher fitness [6]. In terms of multitype branching processes, this can be represented by varying the parameters and offspring probabilities for each type of cell, where cells with higher fitness have a higher probability of splitting, ensuring supercriticality of the process and particular types. Because of the way cells proliferate, we are mostly concerned with binary fission, where each offspring may survive, die, or mutate even though our theory holds for general multitype processes. The transition from normal cells to supercritical cancer cells occurs over multiple replications, and waves of expansion are observed. While the probability of a mutation occurring in a single cell is small, years of replication in a large number of cells makes the likelihood of cancer initiation greater. Sequencing studies have shown that genomes undergo a large number of changes, but most mutations are neutral (the so-called 'passenger mutations') and do not affect cell fitness [2]. In fact, one modeling study determined that half or more somatic mutations occur prior to the cancer initiating event [12]. This means the prior mutations are either passenger mutations, or even if they are driver mutations they do not lead to cells with high enough fitness to overcome normal cells.

The multitype infinite-allele branching process allows modeling of both passenger and driver mutations. We referred to the subpopulations that have different fitness as different types in the model, and to the subpopulations with the same fitness but different genomes as different labels. Cells of different types have different sets of driver mutations in their ancestry, while those with different labels have different sets of passenger mutations. Our results from the model show that the number of mutations grows exponentially and at a rate proportional to the number of individuals alive. Previous studies [2] attempted to determine the correlation between the number of passenger mutations and driver mutations, which we can determine by the number of labels for each type in our model. Alternatively, we can create a more specific process where the type of individual refers to the number of driver mutations present and $K_i(n)$ would represent the total number of passenger alleles associated with i driver mutations. Such a model would allow us to directly compare results to the previous studies. However, the model constrains us to assuming driver mutations are not unique in their effect on growth rates. Our model allows us to get around this constraint and adds more flexibility without requiring the high number of dimensions associated with considering each mutation (driver or passenger) as a specific type.

Acknowledgements

The first author was partially supported by the National Cancer Institute (T32 training grant no. CA096520). The second author was partially supported by the National Science Center (Poland) (grant no. DEC- 2012/04/A/ST7/00353) and the National Science Foundation (grant no. DMS-1361411). We thank Dr Sharon Plon of Baylor College of Medicine for our useful discussions of the applicability of the models and Dr Tony Pakes of The University of Western Australia for providing more details about his work for us to build upon. We also thank the anonymous reviewer for providing us with suggestions to improve the clarity of our work.

References

[1] Athreya, K. B. and Ney, P. E. (1972). *Branching Processes*. Springer, New York.

[2] Bozic, I. *et al.* (2010). Accumulation of driver and passenger mutations during tumor progression. *Proc. Nat. Acad. Sci. USA* **107,** 18545–18550.

[3] Dietzenbacher, E. (1993). A limiting property for the powers of a non-negative, reducible matrix. *Structural Change Econom. Dynamics* **4,** 353–366.
[4] Greenman, C. *et al.* (2007). Patterns of somatic mutation in human cancer genomes. *Nature* **446,** 153–158.
[5] Griffiths, R. C. and Pakes, A. G. (1988). An infinite-alleles version of the simple branching process. *Adv. Appl. Prob.* **20,** 489–524.
[6] Hanahan, D. and Weinberg, R. A. (2000). The hallmarks of cancer. *Cell* **100,** 57–70.
[7] Hoppe, F. M. (1976). Supercritical multitype branching processes. *Ann. Prob.* **4,** 393–401.
[8] Kesten, H. and Stigum, B. P. (1967). Limit theorems for decomposable multi-dimensional Galton–Watson processes. *J. Math. Anal. Appl.* **17,** 309–338.
[9] Kurtz, T., Lyons, R., Pemantle, R. and Peres, Y. (1997). A conceptual proof of the Kesten–Stigum theorem for multi-type branching processes. In *Classical and Modern Branching Processes*, Springer, New York, pp. 181–185.
[10] Mode, C. J. (1971). *Multitype Branching Processes: Theory and Applications*. American Elsevier, New York.
[11] Taïb, Z. (1992). *Branching Processes and Neutral Evolution*. Springer, Berlin.
[12] Tomasetti, C., Vogelstein, B. and Parmigiani, G. (2013). Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc. Nat. Acad. Sci. USA* **110,** 1999–2004.