Article Title

First Author[1,2*], Second Author[2,3†] and Third Author[1,2†]

[1*]Department, Organization, Street, City, 100190, State, Country.
[2]Department, Organization, Street, City, 10587, State, Country.
[3]Department, Organization, Street, City, 610101, State, Country.

*Corresponding author(s). E-mail(s): iauthor@gmail.com;
Contributing authors: iiauthor@gmail.com; iiiauthor@gmail.com;
[†]These authors contributed equally to this work.

Abstract

Stem cells, through their ability to produce daughter stem cells and differentiate into specialized cells, are essential in the growth, maintenance, and repair of biological tissues. Understanding the dynamics of cell populations in the proliferation process not only uncovers proliferative properties of stem cells, but also offers insight into tissue development under both normal conditions and pathological disruption. In this paper, we develop a continuous time branching process model with time-dependent offspring distribution to characterize stem cell proliferation process. We derive analytical expressions for mean, variance, and autocovariance of the stem cell counts, and develop likelihood-based inference procedures to estimate model parameters. Particularly, we construct a forward algorithm likelihood to handle situations when cell count data is partially observed. Simulation results demonstrate that our estimation method recovers the time-dependent division probabilities with good accuracy.

Keywords: branching process, stochastic process, stem cell, cell proliferation, forward algorithm

1 Introduction

Cell proliferation is a fundamental process underlying the growth, maintenance, and repair of biological tissues. Stem cells, in particular, possess the ability to produce identical daughter stem cells and differentiate into specialized cell types that sustain tissue function. There are three types of stem cell division and differentiation

outcomes: symmetrical self-renewal to generate two stem cells, asymmetrical division to generate one stem cell and one differentiated cell, and symmetrical differentiation to generate two differentiated cells. The ability of stem cells to both self-renew and differentiate plays a pivotal role both in generative processes during early development and in tissue regeneration following injury. For example, during embryonic development, stem cells drive tissue formation through programmed proliferation and differentiation, as seen in processes such as ependymogenesis and neurogenesis [1, 2]. Hematopoietic stem cells give rise to all blood cell types - red blood cells, white blood cells, and platelets - to maintain lifelong blood production and regenerate the blood system after injury [3, 4]. In oncology, cancer stem cells also display self-renewal and differentiation ability to drive and sustain tumor growth. Distinct differences in stemness and proliferative potential between cancer stem cells and non-stem cancer cells have been observed across various tumor types such as brain, breast, colon, and prostate cancer [5–8].

It is important to study the dynamics of cell populations in cell proliferation process since these dynamics connect individual cell behaviors to tissue-level outcomes, including how tissues grow, maintain homeostasis, and regenerate after injury. In the aforementioned ependymogenesis process, regional stem cells within the ventricular and subventricular zones proliferate to produce both stem cells and differentiated ependymal cells that line the brain's ventricles, thereby creating a critical barrier between the cerebrospinal fluid and the interstitial fluid of the surrounding brain parenchyma [1]. In the context of cancer study, these population dynamics are particularly important since they reflect the growth of tumor and disease progression.

Quantitative modeling of these processes can provide a framework to predict population behavior and discover intrinsic proliferative properties of stem cell self-renewal and differentiation ability. Deterministic models with ordinary differential equations are utilized to describe the temporal dynamics of cell population on avearge [3, 9–11]. However, since the dynamics of cell populations emerge from the stochastic behavior of individual stem cells, each of which may divide or differentiate at random times, probabilistic modeling approaches, including birth-death process, Gaussian jump model, and non-homogeneous Markov chain [2, 12, 13], are more verisimilar for describing cell proliferation process. Branching process provides a natural and mathematically rigorous framework for describing stochastic reproduction phenomena. In the classical Galton-Watson process and its extensions, each individual acts independently and give rise to a random number of offspring according to specified probability distributions [14–17]. The branching process models have become an important tool to investigate population size, growth and extinction in biology, epidemiology, and genetics. In the context of cell proliferation, branching process models have been applied to describe stem cell lineages, where each stem cell may be viewed as an independent reproducing unit. MacMillian et al. [18] and Miguez [19] both developed models under the branching process theory for the stem cell proliferation and differentiation during embryonic development. In these models, time is discretized using the average cell cycle length as the unit of generations. Instead of tracking individual cell divisions in continuous time, the model advances steps corresponding to the mean duration of a cell cycle. In addition, MacMillian et al. [18] explored the multi-type branching process extension

to simulate data for a cell proliferation with more than one cell type with proliferative potential.

However, the classical branching process models typically assume that the reproduction parameters are constant through time. This assumption simplifies the analysis but may not be able to capture the full stochastic nature of cell proliferation processes. Experimental studies have revealed that the probabilities governing self-renewal and differentiation are not always stationary. For example, data collected by Scanlon et al. [13] on bipotent megakaryocytic erythroid progenitor (MEP) clonal expansion and differentiation shows that probabilities of division outcomes change over time: self-renewing events are more likely earlier in the process and differentiating events are more likely toward the end of the process. The previously discussed branching process framework in Miguez [19] adjusted to the time-dependent proliferation by obtaining the average probabilities of division outcomes at different developmental stages.

In this paper, we develop a continuous-time, time-dependent branching process for stem cell proliferation to address the limitation of the constant reproduction branching process model. In our formulation, the probabilities of division and differentiation outcomes are functions of time. To reflect the variability inherent in stem cell populations, our model incorporates three distinct cell types: viable stem cells capable of division, non-viable stem cells that do not have proliferative potential, and terminally differentiated cells. We also incorporate stochasticity in the timing of the cell division events by assuming that each division occurs at a randomly determined time. In particular, the interarrival times between division events are exponentially distributed with rate proportional to current number of viable stem cells. This reflects the assumption that a greater number of viable stem cells leads to more frequent division events and shorter expected waiting times between them. Our work makes the following contributions. We first establish a rigorous mathematical formulation of the branching process model and derive analytical expressions for key quantities, including mean population size, variance, and extinction time. We also construct likelihood function based on the observed counts of each cell type for the probabilities of division and differentiation outcomes, providing a framework for statistical inference. Finally, we develop a forward-algorithm-based likelihood approach to handle situations where viable and nonviable stem cells cannot be distinguished, allowing for estimation of model parameters even when some cell types cannot be directly observed. This gives a practical modeling approach of stem cell dynamics in experiments where direct observation of all cell types is not possible.

The remainder of this paper is organized as follows. In Section 2, we introduce the time-dependent branching process model and derive analytical expressions for key quantities. Section 3 presents the likelihood-based framework for parameter estimation, including the forward-algorithm approach for partially observed cell types. Section 4 provides simulation results and an illustrative example of experiment data, and Section 5 concludes with a discussion of the findings and potential directions for future work.

2 Time-Dependent Branching Process Model for Stem Cell Proliferation Process

2.1 Model Description

We consider a stochastic cell proliferation process consisting of populations of stem cells and differentiated cells (FC). Stem cells are classified as viable stem cells (SC) and nonviable (or duds) stem cells (DC). Let $X(t), Y(t)$, and $Z(t)$ denote the number of viable stem cells, differentiated cells, and nonviable stem cells at time t, respectively. Only viable stem cells are capable of division and differentiation. The stochastic nature of the process arises from the random timing and outcome of division events. Each viable stem cell divides independently in continuous time with rate r. Biologically, this rate could be interpreted as the average frequency that a viable exits quiescence and enters the cell cycle. Consequently, when there are $X(t)$ viable stem cells at time t, the waiting time until the next division event in the entire population is exponentially distributed with rate $rX(t)$. Thus, the overall frequency of division events increases with the size of the viable stem cell population.

At time t, a viable stem cell can undergo one of the four types of division with the time-dependent probabilities $p_1(t), p_2(t), p_3(t)$, and $p_4(t)$:

Symmetric self-renewal	$SC \to SC + SC$,	probability $p_1(t)$
Asymmetric division	$SC \to SC + FC$,	probability $p_2(t)$
Symmetric differentiation	$SC \to FC + FC$,	probability $p_3(t)$
Self renewal with nonviable stem cell	$SC \to SC + DC$,	probability $p_4(t)$

The division probabilities $p_j(t)$ are time-dependent. For example, $p_1(t)$ may dominate early to expand the stem cell pool, while $p_3(t)$ may increase toward the end of the cell proliferation process. Since $p_j(t)$ are probabilities, they need to satisfy the conditions $0 \le p_j(t) \le 1$ and $\sum_{j=1}^{n} p_j(t) = 1$ $\forall t$. To illustrate, we may assume the following form for $p_j(t)$

$$\begin{cases} p_1(t) = \frac{p_1}{1+c_1(t-m_1)^2}, \\ p_2(t) = \frac{p_2}{1+c_2(t-m_2)^2}, \\ p_3(t) = 1 - \frac{p_1}{1+c_1(t-m_1)^2} - \frac{p_2}{1+c_2(t-m_2)^2} - \frac{p_4}{1+c_4(t-m_4)^2}, \\ p_4(t) = \frac{p_4}{1+c_4(t-m_4)^2}, \end{cases} \tag{1}$$

with the constraints, $0 \le p_1, p_2, p_4 \le 1$, $p_1 + p_2 + p_4 \le 1$, and $c_1, c_2, c_4 \ge 0$. In this example, the functions $p_1(t), p_2(t)$, and $p_4(t)$ are Lorentzian functions that are parameterized by p_i, which are the maximum of the function attained at $t = m_i$, and c_i, which governs the rate of decay away from the maximum. With this form of functions for $p_j(t)$, $p_3(t) \to 1$ as $t \to \infty$. As a result, the differentiation events will dominate toward the end of the proliferation process. This behavior aligns with the dynamics observed in the MEP clonal expansion and differentiation data observed by Scanlon et al. [13].

This proliferation process can be naturally represented as a branching process, in which each viable stem cell acts as a branching individual producing "offspring"

according to the division probabilities $p_j(t)$. Differentiated cells correspond to terminal offspring, and nonviable stem cells correspond to individuals that cannot further proliferate. Thus, given the sequence of event times $T_1, T_2, \cdots$, the change of viable stem cells, differentiated cells, and non-viable stem cell have the following marginal distributions

$$P(\Delta X_i, \Delta Y_i, \Delta Z_i | T_i) = \begin{cases} p_1(T_i), & (\Delta X_i, \Delta Y_i, \Delta Z_i) = (1, 0, 0) \\ p_2(T_i), & (\Delta X_i, \Delta Y_i, \Delta Z_i) = (0, 1, 0) \\ p_3(T_i), & (\Delta X_i, \Delta Y_i, \Delta Z_i) = (-1, 2, 0) \\ p_4(T_i), & (\Delta X_i, \Delta Y_i, \Delta Z_i) = (0, 0, 1). \end{cases} \tag{2}$$

A deterministic mean-field model can capture the average behavior of the cell population by assuming individual viable stem cells reproduce at their average rate. Let $S(t), F(t), D(t)$ denote the deterministic population trajectories of viable stem cells, differentiated cells, and non-viable stem cells. If we assume that the average rates at which viable stem cells, differentiated cells, and non-viable stem cells are reproduced are $r[p_1(t) - p_3(t)]S(t)$, $r[p_2(t) + 2p_3(t)]S(t)$, and $rp_4(t)S(t)$, respectively, we obtain the ordinary differential equation

$$\begin{pmatrix} \frac{dS(t)}{dt} \\ \frac{dD(t)}{dt} \\ \frac{dF(t))}{dt} \end{pmatrix} = \begin{pmatrix} r[p_1(t) - p_3(t)]S(t) \\ rp_4(t)S(t) \\ r[p_2(t) + 2p_3(t)]S(t) \end{pmatrix}. \tag{3}$$

A limitation of the deterministic model is that it treats the probabilities of division and differentiation events as average rates, and therefore does not capture the randomness of the underlying process. As a result, the deterministic model cannot account for the variability between realizations or extinction events.

2.2 Mean and Covariance Structure of Viable Stem Cell Population

In this subsection, we derive the expected value of the viable stem cell population in the branching process model and show that it coincides with the solution of the deterministic differential equation. We also compute the variance and autocovariance, which measure the variability of the population at a single time point and across time points, respectively.

Proposition 1 (Moments of Number of Stem Cells) *For each fixed time point, the expected value, variance, and autocovariance of the stem cell population are given by:*

*1. **Expected Value (Mean Function).***

$$\mathbb{E}[X(t)] = S(t) = S_0 \exp\left\{ r \int_0^t [p_1(u) - p_3(u)du] \right\}.$$

2. **Variance.**
$$\text{Var}(X(t)) = rS(t)^2 \int_0^t \frac{p_1(u) + p_3(u)}{S(u)} \, du.$$

3. **Autocovariance Function.** *For $t > u$,*

$$\text{Cov}(X(t), X(u)) = S(t)S(u)r \int_0^u \frac{p_1(v) + p_3(v)}{S(v)} \, dv = \frac{S(t)}{S(u)} \text{Var}(u).$$

Proof The solution of the differential equation (3) with initial condition $S(0) = S_0$ is given by

$$S(t) = S_0 \exp\left\{ r \int_0^t \left[p_1(u) - p_3(u) \right] du \right\}. \tag{4}$$

We first show that $E[X(t)] = S(t)$. For $\Delta > 0$, conditional on $X(t)$,

$$\begin{aligned} E[X(t + \Delta) - X(t) \mid X(t)] &= \left[p_1(t) - p_3(t) + O(\Delta) \right] rX(t)\Delta \\ &= \left[p_1(t) - p_3(t) \right] rX(t)\Delta + o(\Delta). \end{aligned} \tag{5}$$

Taking expectations yields

$$E[X(t + \Delta)] - E[X(t)] = \left[p_1(t) - p_3(t) \right] rE[X(t)]\Delta + o(\Delta). \tag{6}$$

Dividing by Δ and letting $\Delta \to 0$ gives

$$\frac{d}{dt} E[X(t)] = \left[p_1(t) - p_3(t) \right] rE[X(t)]. \tag{7}$$

Since $E[X(0)] = S_0$, uniqueness of solutions to linear ODEs implies $E[X(t)] = S(t)$.

We next derive the variance of $X(t)$. Define $M(t) = E[X(t)^2]$ and $V(t) = \text{Var}(X(t)) = M(t) - S(t)^2$. Conditioning on $X(t)$,

$$E\left[X(t + \Delta)^2 - X(t)^2 \mid X(t) \right] = \left[2X(t)(p_1(t) - p_3(t)) + (p_1(t) + p_3(t)) \right] rX(t)\Delta. \tag{8}$$

Taking expectations yields

$$M(t + \Delta) - M(t) = 2\left[p_1(t) - p_3(t) \right] rM(t)\Delta + \left[p_1(t) + p_3(t) \right] rS(t)\Delta. \tag{9}$$

Letting $\Delta \to 0$ gives

$$M'(t) = 2\left[p_1(t) - p_3(t) \right] rM(t) + \left[p_1(t) + p_3(t) \right] rS(t). \tag{10}$$

Since $S'(t) = \left[p_1(t) - p_3(t) \right] rS(t)$, differentiation of $V(t) = M(t) - S(t)^2$ yields

$$V'(t) = 2\left[p_1(t) - p_3(t) \right] rV(t) + \left[p_1(t) + p_3(t) \right] rS(t). \tag{11}$$

Assuming $V(0) = 0$, the solution of (11) is

$$V(t) = rS(t)^2 \int_0^t \frac{p_1(u) + p_3(u)}{S(u)} \, du, \tag{12}$$

where $P(t) = \int_0^t [p_1(u) - p_3(u)] \, du$. Finally, for $t > u$, the tower property and conditional expectation give

$$\begin{aligned} E[X(t)X(u)] &= E[E[X(t) \mid X(u)] X(u)] \\ &= E\left[X(u)^2 \exp\{r[P(t) - P(u)]\} \right] \\ &= \frac{S(t)}{S(u)} E[X(u)^2]. \end{aligned} \tag{13}$$

Using (12) with $M(u) = S(u)^2 + V(u)$, we obtain

$$E[X(t)X(u)] - S(t)S(u) = \frac{S(t)}{S(u)} V(u). \tag{14}$$

This completes the proof. □

We validate the theoretical moments of stem cell count using simulated data. Details on how the data is simulated is discussed in section 4. Figure 1 (left) and figure 2 illustrate an example of empirical mean, variance and autocorrelation of stem cell counts obtained from simulated data. The empirical results closely match with their theoretical counterparts derived in proposition 1.
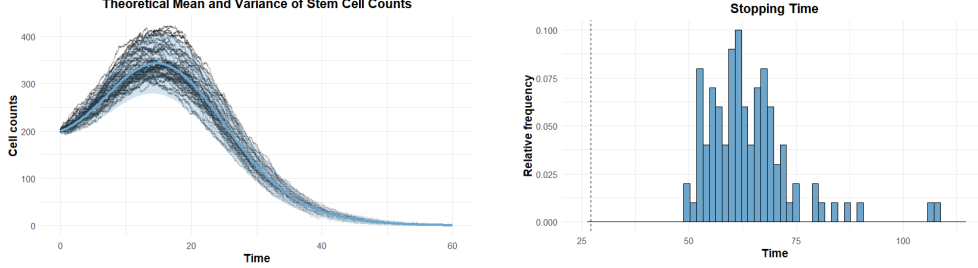

Fig. 1 Left: Fifty realizations of simulated stem cell count trajectories with initial count of 200, division rate of 0.2, and division probabilities $p_1(t) = \frac{0.55}{1+0.005(t-4)^2}$, $p_2(t) = \frac{0.15}{1+(t-12)^2}$, $p_4(t) = \frac{0.20}{1+(t-20)^2}$, and $p_3(t) = 1 - p_1(t) - p_2(t) - p_4(t)$. The blue curve shows the expected cell count and the shaded band indicates ± 2 standard deviations. Right: Distribution of stopping time, i.e. the time when all the viable stem cells are exhausted from 100 realizations of the simulated stem cell count trajectories with the same parameters, with the minimum expected stopping times is $\tau_{\min} = 26.491$.

<div align="center">

Theoretical autocorrelation   **Empirical autocorrelation**

</div>

t	5	10	15	20	25
5	1.000	0.763	0.666	0.598	0.533
10	0.763	1.000	0.872	0.783	0.698
15	0.666	0.872	1.000	0.899	0.801
20	0.598	0.783	0.899	1.000	0.891
25	0.533	0.698	0.801	0.891	1.000

t	5	10	15	20	25
5	1.000	0.742	0.588	0.552	0.535
10	0.742	1.000	0.884	0.761	0.758
15	0.588	0.884	1.000	0.877	0.849
20	0.552	0.761	0.877	1.000	0.891
25	0.535	0.758	0.849	0.891	1.000

Fig. 2 Theoretical versus empirical autocorrelation of stem cell counts at time points $t = 5, 10, 15, 20, 25$. The empirical autocorrelation is calculated from realizations of simulated stem cell trajectories with initial count of 200, division rate of 0.2, and division probabilities $p_1(t) = \frac{0.55}{1+0.005(t-4)^2}$, $p_2(t) = \frac{0.15}{1+(t-12)^2}$, $p_4(t) = \frac{0.20}{1+(t-20)^2}$, and $p_3(t) = 1 - p_1(t) - p_2(t) - p_4(t)$.

Remark 1 Suppose that

$$P(t) = \int_0^t \left[p_1(u) - p_3(u) \right] du \longrightarrow -\infty \qquad \text{as } t \to \infty.$$

Then both the mean and the variance of $X(t)$ converge to zero.

The convergence $E[X(t)] \to 0$ follows immediately from the explicit expression

$$E[X(t)] = S_0 \exp\{rP(t)\}.$$

To analyze the variance, recall that

$$V(t) = rS_0 \exp\{2rP(t)\} \int_0^t \exp\{-rP(u)\} \left[p_1(u) + p_3(u) \right] du.$$

7

Since $0 \le p_1(t) + p_3(t) \le 1$, we obtain the bound

$$V(t) \le rS_0 \exp\{2rP(t)\} \int_0^t \exp\{-rP(u)\}\, du. \tag{15}$$

Applying l'Hôpital's rule to the ratio

$$\frac{\int_0^t \exp\{-rP(u)\}\, du}{\exp\{-2rP(t)\}},$$

we find

$$\lim_{t\to\infty} \frac{\int_0^t \exp\{-rP(u)\}\, du}{\exp\{-2rP(t)\}} = \lim_{t\to\infty} \frac{\exp\{rP(t)\}}{-2r\big[p_1(t) - p_3(t)\big]}.$$

Under the assumption $P(t) \to -\infty$, the numerator converges to zero, while the denominator remains bounded away from zero whenever $p_1(t) - p_3(t)$ does not vanish identically. Consequently, the right-hand side of (15) converges to zero, and hence

$$\lim_{t\to\infty} V(t) = 0.$$

Intuitively, this result is expected because over time, as differentiation events dominate self-renewal events, stem cells are lost faster than they are replenished and become extinct. Moreover, when the initial number of stem cells is large, for any fixed time point t the stochastic process is close to their expected values. Specifically, we have the following limit result.

Proposition 2 *For fixed t, as $S_0 \to \infty$,*

$$\frac{X(t) - S(t)}{S(t)} \xrightarrow{L_2} 0. \tag{16}$$

Proof Since we have $E[X(t)] = S(t)$, we only need to show $\mathrm{Var}\left(\frac{X(t)-S(t)}{S(t)}\right) \longrightarrow 0$ as $S_0 \longrightarrow \infty$.

$$\mathrm{Var}\left(\frac{X(t) - S(t)}{S(t)}\right) = \frac{r}{S_0} \int_0^t \frac{p_1(u) + p_3(u)}{\exp\{rP(u)\}} du \longrightarrow 0. \tag{17}$$

$\square$

This result implies that, when starting with a large number of initial stem cells, the stem-cell population in the branching process can be well approximated by the deterministic model with the differential equation in (3).

2.3 Stopping Time

Since viable stem cells are the only cell type with proliferative capability in our model, the process terminates when they become extinct. In this section, we analyze the stopping time of the process, defined as the time at which the number of viable stem cells reach 0.

We first consider the lower bound of the stopping times. Under the model assumption that interarrival times between two consecutive division events is the a exponential distribution with rate proportional to the current viable stem cells, the

average interarrival time between two consecutive division events $T_{i+1} - T_i$ is $1/(rS_i)$. The minimum stopping times occurs when every event is a differentiation event, in which no new viable stem cell is created and the number of viable stem cells decreases by 1. Then the minimum expected stopping time is

$$\tau_{\min} = \frac{1}{rS_0} + \frac{1}{r(S_0 - 1)} + \cdots + \frac{1}{r}.$$

Using the harmonic sum approximation,

$$\tau_{\min} \approx (1/r)\log(S_0) + \gamma, \tag{18}$$

where $\gamma \approx 0.5772$ is the Euler–Mascheroni constant. Note that this is a conservative lower bound of the stopping time since we assume that every division results in a loss of one stem cell (i.e. $p_3(t) = 1 \ \forall t$). Figure 1 (right) presents an example of stopping time distribution from simulated data. The minimum expected stopping time lies below all simulated stopping times.

We next fit the stopping times from simulated data using the inverse Gaussian distribution. The inverse Gaussian distribution is commonly used to model first passage or stopping times of processes such as Brownian motion with drift [20]. In our context, the inverse Gaussian distribution provides a natural model for the time it takes for the number of viable stem cells to reach 0. The simulated stopping data is obtained from 100 realizations of the process with the initial count of 200, division rate of 0.2, and division probabilities $p_1(t) = \frac{0.55}{1+0.005(t-4)^2}$, $p_2(t) = \frac{0.15}{1+(t-12)^2}$, $p_4(t) = \frac{0.20}{1+(t-20)^2}$, and $p_3(t) = 1 - p_1(t) - p_2(t) - p_4(t)$. Since the support of the inverse Gaussian is the positive real line, we shift the stopping times by $-\tau_{\min}$. Using these data, we estimate the mean and shape parameter of the inverse Gaussian distribution by the maximum likelihood method. The estimated parameters are then used to fit the distribution to the simulated data, and the goodness-of-fit is assessed through standard distributional tests. The results of distributional tests in table 1 and diagnostic plots in figure 5 in the Supplementary Materials indicate that the inverse Gaussian distribution is a good fit for the shifted simulated stopping times.

Distribution test	Test statistics	P-value
Kolmogorov-Smirnov	0.066	0.766
Anderson-Darling	0.588	0.658

Table 1 Distributional tests of the shifted stopping times with the inverse Gaussian distribution.

3 Likelihood Estimation

In practice, the division probabilities $p_1(t), p_2(t), p_3(t), p_4(t)$ are unknown and must be estimated using the observed cell proliferation data from experiments. In this section,

we develop a likelihood-based framework for parameter estimation. When we observe all the events, both the timing of the events and the number of cells, we can write down the full likelihood of the observed data

$$\mathbb{L}(\boldsymbol{\theta}; \mathbb{T}, \mathbb{X}, \mathbb{Y}, \mathbb{Z}) = \prod_{i=1}^{n} P(\Delta X_i, \Delta Y_i, \Delta Z_i | T_i) f(\Delta T_i | X_{i-1}), \tag{19}$$

where θ denotes the set of parameters for $p_j(t)$ and division rate r, $P(\Delta X_i, \Delta Y_i, \Delta Z_i | T_i)$ is the distribution of the change of cell counts in (2), and $f(\cdot | X_{i-1})$ denotes the density function of the exponential distribution with rate $r X_{i-1}$. Note that in the likelihood for fully observed data, we only need to use the observed viable stem cells and non-viable stem cells to determine which division event occurs. The likelihood can be optimized to obtain the maximum likelihood estimates of $p_i(t)$ parameters and r. In practical experiments, we may encouter situations in which viable and non-viable stem cells cannot be distinguished. As a result, instead of observing $X(t)$ and $Z(t)$, we observe the sum of the viable and non-viable stem cell denoted by $M(t)$. As seen in (19), the unobserved number of nonviable stem cells plays an important role in exact likelihood evaluation. A naïve approach to evaluate the likelihood when the nonviable stem cells are not distinguishable from viable stem cells is to sum out all the trajectories of possible numbers of nonviable stem cells, which is computationally inefficient. In such cases, we employ a modification of the forward algorithm, which is a dynamic programming technique developed for hidden Markov models [21], to write the likelihood of partially observed data.

We first define the forward variables $\alpha_k(u)$ (for $k = 1, \cdots, n$) as the likelihood that there are u unobserved non-viable stem cells at step k, given all observations (M_i, Y_i, T_i) up to and including step k

$$\alpha_k(u) = \mathbb{L}(\mathbb{T}_k, \mathbb{Y}_k, \mathbb{M}_k, Z_k = u, \boldsymbol{\theta}),$$

where u is an integer and $0 \le u \le M_k$, $\mathbb{T}_k, \mathbb{Y}_k, \mathbb{M}_k$ are the observations of event times, differentiated cells, and total stem cells up to step k, respectively. The forward variables can be computed recursively by

$$\alpha_{k+1}(v) = \sum_{u \in \{v, v-1\}} \alpha_k(u) \cdot h_k(u, v), \quad 0 \le k \le n-1$$

where $0 \le v \le M_{k+1}$. Note that since we assume $T_0 = 0, M_0 = S_0, Y_0 = 0, Z_0 = 0$, we have $\alpha_0(0) = 1$ and $\alpha_0(u) = 0$ for $u > 0$. Since the number of nonviable stem cell either stay the same or increase by 1 after each division, we only need to sum over these two cases. Here, $h_k(u, v)$ are analogous to transitional probabilities from having u unobserved non-viable stem cells at step k to having v unobserved non-viable stem cells at step $k + 1$. For $i \le j \le 4$

$$h_k(u, v) = \begin{cases} h_{jk}(u, v), & \text{event } j \text{ at } T_{k+1} \\ 0, & \text{otherwise,} \end{cases}$$

10

$h_{jk}(u,v) = r(M_k - u)e^{-r(M_k-u)(T_{k+1}-T_k)} \cdot p_j(T_{k+1})$. The events in terms of (Y, M, Z) are:

- Event 1: $Y_{k+1} = Y_k, M_{k+1} = M_k + 1, v = u$ with probability $p_1(T_{k+1})$,
- Event 2: $Y_{k+1} = Y_k + 1, M_{k+1} = M_k, v = u$ with probability $p_2(T_{k+1})$,
- Event 3: $Y_{k+1} = Y_k + 2, M_{k+1} = M_k - 1, v = u$ with probability $p_3(T_{k+1})$,
- Event 4: $Y_{k+1} = Y_k, M_{k+1} = M_k + 1, v = u + 1$ with probability $p_4(T_{k+1})$.

The likelihood is given by

$$\mathbb{L}_n = \sum_{0 \leq u \leq M_n} \alpha_n(u).$$

To avoid the issue of underflow or overflow in computation, we apply a standard normalization technique to the forward algorithm. The normalized forward variable is defined as

$$\bar{\alpha}_k(u) = \frac{\alpha_k(u)}{\mathbb{L}_k},$$

where $\mathbb{L}_k = \sum_{0 \leq u \leq M_k} \alpha_k(u), \ k = 1, \cdots, n$. The recursion of the normalized forward variable is then given by

$$\bar{\alpha}_{k+1}(v) = \frac{\alpha_{k+1}(v)}{\mathbb{L}_{k+1}} = \frac{1}{\mathbb{L}_{k+1}} \sum_{u=v-1,v} \alpha_k(u)h_k(u,v) = \frac{\mathbb{L}_k}{\mathbb{L}_{k+1}} \sum_{u=v-1,v} \bar{\alpha}_k h_k(u,v). \quad (20)$$

For $0 \leq k \leq n - 1$, we define the sequential likelihood ratio

$$
\begin{aligned}
d_{k+1} &= \frac{\mathbb{L}_{k+1}}{\mathbb{L}_k} \\
&= \frac{1}{\mathbb{L}_k} \sum_{0 \leq v \leq M_{k+1}} \alpha_{k+1}(v) \\
&= \frac{1}{\mathbb{L}_k} \sum_{0 \leq v \leq M_{k+1}} \sum_{u=v-1,v} \alpha_k(u)h_k(u,v) \\
&= \sum_{0 \leq v \leq M_{k+1}} \sum_{u=v-1,v} \bar{\alpha}_k(u)h_k(u,v)
\end{aligned}
\quad (21)
$$

The log-likelihood of the observed data is obtained by taking the logarithm of the sequential likelihood ratios, and performing the telescopic summation

$$\log(\mathbb{L}_n) = \sum_{k=0}^{n-1} \log(d_{k+1}).$$

Algorithm (1) provides a formal description of the normalized forward algorithm. It can be used to optimize and obtain the maximum likelihood estimation. The results of the algorithm implementation and parameter estimation are presented in the next section.

Algorithm 1 Normalized Forward Algorithm to Compute Log-likelihood ($\log \mathbb{L}_n$)

Require: Parameter of probabilities $p_j(t)$ and division rate r, observed event times $\mathbb{T} = (t_1, \cdots, t_n)$, observed total stem cells $\mathbb{M} = (m_1, \cdots m_n)$ and differentiated cells $\mathbb{Y} = (y_1, \cdots, y_n)$.

Ensure: Log-likelihood $\log \mathbb{L}_n$.

    Initialization:

1:   $\bar{\alpha}_0(0) = 1$

    Recursion:

2:   **for** $k = 0$ to $n - 1$ **do**

3:      **for** $u \in \{v - 1, v\}$, $0 \leq v \leq M_{k+1}$ **do**

4:        $h_k(u, v)$

5:      **end for**

6:      $d_{k+1} = \displaystyle\sum_{0 \leq v \leq M_{k+1}} \sum_{u = v-1, v} \bar{\alpha}_k(u) h_k(u, v)$

7:      **for** $u \in \{v - 1, v\}$ **do**

8:        $\bar{\alpha}_{k+1}(v) = \frac{1}{d_{k+1}} \displaystyle\sum_{u} \bar{\alpha}_k(u) h_k(u, v)$

9:      **end for**

10: **end for**

    Termination:

11: $\log \mathbb{L}_n = \displaystyle\sum_{k=0}^{n-1} \log d_{k+1}$

12: **return** $\log \mathbb{L}_n$

---

4 Simulation Results

In this section, we assess the performance of the maximum likelihood estimator (MLE) for the branching process model of the stem cell proliferation process. We consider both fully observed cell count data and scenarios in which only partial observations are available. Simulated cell count trajectories are generated as follows. Starting from a fixed number of viable stem cells, the time until the next event was drawn from an exponential distribution with rate proportional to the current number of viable stem cells. At each event, a stem cell either divides or differentiates according to probabilities $p_j(t)$. In our simulation, the probabilities $p_j(t)$ are time-dependent with the functional form described in (1). After each event, the population of each cell type is updated, and the process is repeated until all viable stem cells are exhausted. For partially observed data, we do not observe the counts of viable and non-viable stem cells separately, but only observe their combined total. In our estimation method, we only need to track the timing of the events and changes of cell counts after each division event rather than the full lineage of individual cells. We do not need to record when each cell is created or undergoes division. Consequently, our estimation method reduces the need for extensive experimental data collection. We explore three parameter configurations to evaluate how differences in maximum and timing affect the overall probability dynamics of the division and differentiation events. Figure 3 displays the
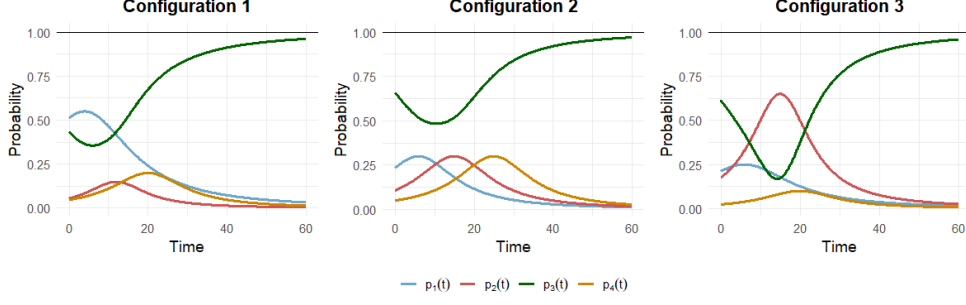
Fig. 3 Time-depedent probability functions for division and differentiation events under three Lorentzian parameter configurations. Each panel displays the four probabilities $p_1(t), p_2(t), p_3(t)$, and $p_4(t)$ defined in equation (1).

plots of probabilities $p_j(t)$, $j = 1, 2, 3, 4$ with these three different parameter configurations. Configuration 1 represents an early dominance of the symmetric self-renewal divisions with a large initial peak in $p_1(t)$. Configuration 2 provides a balanced case in which functions $p_1(t), p_2(t), p_4(t)$ have the equal maximum and rate of decay, but the maximums occur at well-separated times. Configuration 3 emphasizes on a strong peak in $p_2(t)$, which is the asymmetric division into one stem cell and one differentiated cell. For each parameter configuration, 100 replicate trajectories are simulated. The number of initial viable stem cells and division rate used in the simulation are 200 and 0.2, respectively. The likelihood of fully observed data is computed based on equation (19). For numerical stability, we take the logarithm of the likelihood of fully observed data

$$\ell(\boldsymbol{\theta}, \mathbb{T}, \mathbb{X}, \mathbb{Y}, \mathbb{Z})$$

$$= \sum_{i=1}^{k} \Bigg[\log \left(\frac{p_1}{1 + c_1(T_i - m_1)^2} \right) I_{(\Delta X_i = 1, \Delta Z_i = 0)}$$

$$+ \log \left(\frac{p_2}{1 + c_2(T_i - m_2)^2} \right) I_{(\Delta X_i = 0, \Delta Z_i = 0)}$$

$$+ \log \left(\frac{p_4}{1 + c_4(T_i - m_4)^2} \right) I_{(\Delta X_i = 0, \Delta Z_i = 1)}$$

$$+ \log \left(1 - \frac{p_1}{1 + c_1(T_i - m_1)^2} - \frac{p_2}{1 + c_2(T_i - m_2)^2} - \frac{p_4}{1 + c_4(T_i - m_4)^2} \right) I_{(\Delta X_i = -1, \Delta Z_i = 0)}$$

$$+ \log r + \log S_{i-1} - r S_{i-1} \Delta_{T_1} \Bigg].$$

The closed-form solution of the MLE for division rate r can be obtained by taking the derivative of the log-likelihood with respect to r. The likelihood of the partially observed cell counts is computed using the forward algorithm described in section 3. Parameters are estimated by numerical maximization of the likelihood: for fully observed data, we use the differential evolution (DE) algorithm for parameters of

$p_j(t)$, while the MLE of division rate can be obtained through its closed-form solution. DE is a global optimization algorithm which maintains a population of candidate solutions and iteratively improves them through mutation, crossover, and selection. In the mutation step, new candidate solutions are generated by adding a scaled difference between two population members to a third member, allowing the algorithm to explore the parameter space efficiently. These candidates are then mixed with existing solutions through crossover to increase diversity of the population of candidates, with the best-performing candidates are selected to retain for the next generation. DE is particularly well-suited for complex or non-convex likelihood surfaces, as it does not require gradient information and is less likely to be trapped in local optima [22–24]. For partially observed data, DE is combined with the gradient-based Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization, with the gradient computed by automatic differentiation [25, 26] to refine estimates near local maxima. This hybrid approach takes advantage of the global search capability of DE to explore the parameter space broadly, while BFGS with automatic differentiation refines solutions locally, improving convergence speed and numerical stability. Note that although we employ automatic differentiation to evaluate the gradient of the normalized forward-algorithm-based likelihood, the gradient can also be computed dynamically. The full derivation is provided in the Supplementary Materials.

For the fully observed data case (Table 2), the means and medians of the estimates closely match the true parameter values and the 5th-95th percentile intervals are narrow across all three configurations. Table 3 summarizes the estimates obtained with the normalized forward algorithm for scenarios in which we cannot distinguish viable and nonviable stem cells. The result shows a slight increase in the variability of the estimates as the percentile ranges are slightly wider. The true parameters are still contained within the corresponding intervals for all parameters. Across three configurations, the location parameters m_i are estimated particularly well, whereas the estimates of parameters c_i experience higher degree of sensitivity to loss of information due to unobserved data. We assess the fit of the model by comparing the expected cell count trajectories generated from the estimated parameters from the partially observed data and forward algorithm likelihood with the corresponding observed simulated data. We use the ratio between the observed and expected cell counts

$$\frac{X(t)}{\hat{X}(t)}, \frac{Z(t)}{\hat{Z}(t)}, \frac{M(t)}{\hat{M}(t)}, \frac{Y(t)}{\hat{Y}(t)},$$

goodness-of-fit assessment metrics. If the ratios are close to 1, the model provides a good estimation of the trajectory of cell counts. Figure 4 depicts the observed-to-expected cell count ratios across four panels: viable stem cells, nonviable stem cells, total stem cells, and differentiated cells. The ratios for differentiated cells are tightly clustered around one, indicating a good fit of the model to the observed data. Total stem cell ratios are also well reproduced. The ratios for nonviable stem cells show the greatest variability, especially at early time points. Overall, these results suggest that the estimation of the model parameters accurately captures the dynamics of differentiated and total stem cells, with larger discrepancies primarily occurring for nonviable stem cells during the initial phase.

14

Parameter	True value	Mean	Median	5.0 Percentile	95.0 Percentile
Parameter configuration 1					
p_1	0.550	0.560	0.557	0.524	0.602
p_2	0.150	0.151	0.148	0.123	0.181
p_4	0.200	0.201	0.200	0.175	0.242
c_1	0.005	0.00504	0.00491	0.00365	0.00704
c_2	0.012	0.01270	0.01170	0.00653	0.02050
c_4	0.008	0.00854	0.00826	0.00543	0.01280
m_1	4	3.500	3.740	0.549	5.750
m_2	12	11.900	11.900	10.100	13.500
m_4	20	19.900	20.000	18.000	21.800
r	0.200	0.199	0.199	0.191	0.208
Parameter configuration 2					
p_1	0.300	0.306	0.303	0.256	0.373
p_2	0.300	0.304	0.303	0.250	0.354
p_4	0.300	0.306	0.305	0.246	0.372
c_1	0.008	0.00894	0.00790	0.00389	0.01690
c_2	0.008	0.00872	0.00800	0.00465	0.01390
c_4	0.008	0.00882	0.00831	0.00442	0.01430
m_1	6	5.650	5.920	2.21	8.29
m_2	15	15.100	15.200	13.300	17.2
m_4	25	24.900	25.000	22.000	28.500
r	0.200	0.199	0.199	0.186	0.211
Parameter configuration 3					
p_1	0.250	0.258	0.258	0.223	0.295
p_2	0.650	0.649	0.652	0.585	0.709
p_4	0.100	0.105	0.102	0.0713	0.147
c_1	0.005	0.00560	0.00526	0.00272	0.00938
c_2	0.012	0.01220	0.01220	0.00887	0.01590
c_4	0.008	0.01070	0.00927	0.00326	0.02130
m_1	6	5.590	6.130	0.501	9.290
m_2	15	15.000	15.000	14.200	15.700
m_4	20	19.500	19.500	16.000	23.400
r	0.200	0.199	0.198	0.189	0.210

Table 2 Parameter estimations under fully observed data for the three parameter configurations.

5 Discussion

In this work, we have formulated a continuous-time branching process model for cell proliferation with time-dependent division probabilities. Unlike classical branching models with fixed offspring distribution, our framework accommodates proliferation processes in which the cell fate decisions may evolve during development.

One key contribution of this work is the development of the likelihood-based inference procedures under two different observation scenarios. When all division events and cell counts are observable, we derived a full likelihood that allows direct parameter estimation. More importantly, when cell counts are partially unobservable - viable and non-vaible stem cells cannot be distinguished - we constructed a forward algorithm likelihood. This approach avoids the need to calculate the likelihood from all possible divisions resulting in the observed cell counts, greatly reducing the complexity of the

Parameter	True value	Mean	Median	5.0 Percentile	95.0 Percentile
Parameter configuration 1					
p_1	0.550	0.565	0.563	0.512	0.621
p_2	0.150	0.151	0.149	0.121	0.182
p_4	0.200	0.210	0.205	0.170	0.261
c_1	0.005	0.00539	0.00513	0.00286	0.00875
c_2	0.012	0.01280	0.01180	0.00617	0.02070
c_4	0.008	0.01030	0.00959	0.00447	0.01920
m_1	4	3.510	3.700	0.01850	6.070
m_2	12	11.900	12.000	10.000	13.500
m_4	20	19.900	20.100	15.700	22.900
r	0.200	0.198	0.197	0.187	0.213
Parameter configuration 2					
p_1	0.300	0.313	0.315	0.246	0.375
p_2	0.300	0.304	0.305	0.247	0.355
p_4	0.300	0.311	0.307	0.223	0.409
c_1	0.008	0.00979	0.00898	0.00267	0.02000
c_2	0.008	0.00879	0.00808	0.00446	0.01460
c_4	0.008	0.01000	0.00892	0.00351	0.01910
m_1	6	5.610	6.010	0.184	9.030
m_2	15	15.100	15.100	13.200	17.200
m_4	25	24.900	25.000	21.200	28.800
r	0.200	0.197	0.196	0.183	0.218
Parameter configuration 3					
p_1	0.250	0.271	0.275	0.208	0.329
p_2	0.650	0.658	0.662	0.587	0.714
p_4	0.100	0.135	0.129	0.08750	0.207
c_1	0.005	0.00746	0.00711	0.00179	0.01460
c_2	0.012	0.01260	0.01230	0.00906	0.01680
c_4	0.008	0.02440	0.01640	0.00314	0.06360
m_1	6	6.610	6.480	0.227	15.100
m_2	15	15.000	15.000	14.200	15.800
m_4	20	18.500	21.000	0.294	26.800
r	0.200	0.198	0.196	0.183	0.224

Table 3 Parameter estimation when viable and nonviable stem cells cannot be distinguised using normalized forward algorithm and hybrid optimization.

likelihood computation. Simulation studies demonstrate that the proposed estimation recovers the time-dependent division probabilities with good accuracy under both scenarios from the aggregate cell count data and the timing of the division events. In both scenarios, our method does not require to track individual cell lineage, which are often difficult to obtain in experiments. Our results suggest that meaningful information about proliferation probabilities can still be recovered where lineage tracking is not feasible. From an experimental perspective, this substantially reduces the work required to track and collect data.

Our proposed framework can be extended in useful ways to study more complex cell proliferation processes. The number of division outcomes to can be expanded to include additional progenitor or differentiated cell types, leading to higher-dimensional multi-type branching processes. In addition, while the current model allows division
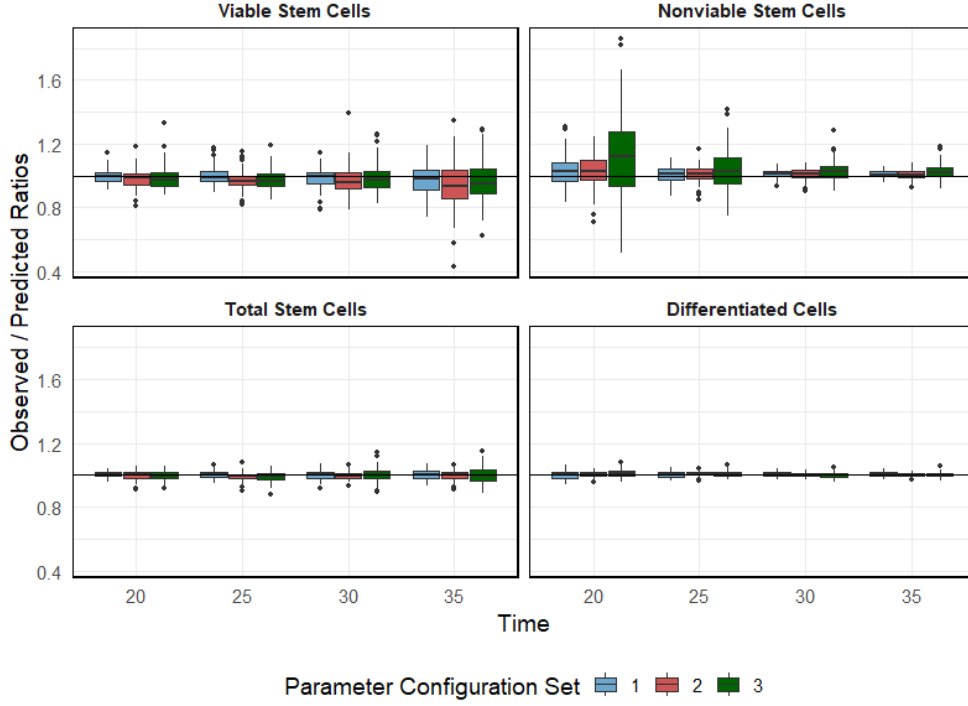
Fig. 4 Ratios of observed-to-expected cell counts. The expected cell counts are generated by using the parameter estimates for partially observed data with forward algorithm likelihood in table 3

probabilities to be time-dependent, they can be generalized to depend on other covariates, such as spatial location or local cell density. We expect that our approach will be useful for analyzing a wide range of biological processes in which cell fate decisions evolve dynamically and are only partially observable.

Supplementary Materials

Fitting Inverse Gaussian Distribution to the Stopping Time

The simulated stopping data is obtained from 100 realizations of the process with the initial count of 200, division rate of 0.2, and division probabilities $p_1(t) = \frac{0.55}{1+0.005(t-4)^2}$, $p_2(t) = \frac{0.15}{1+(t-12)^2}$, $p_4(t) = \frac{0.20}{1+(t-20)^2}$, and $p_3(t) = 1-p_1(t)-p_2(t)-p_4(t)$. The simulated stopping time data is then shifted by $-\tau_{\min}$. The maximum likelihood estimates of mean and shape parameter of the inverse Gaussian distribution are shown in table 4. The estimates are then used to fit the distribution to the simulated data.

Parameter	Mean	Shape
Estimates	36.858	586.698

Table 4 Maximum likelihood estimates of inverse Gaussian distribution parameters with the shifted simulated stopping time.

Gradient of the Forward Likelihood

For $1 \leq v \leq M_{k+1}$, the recursion of the normalized forward variable is given by

$$\bar{\alpha}_{k+1}(v) = \frac{\mathbb{L}_k}{\mathbb{L}_{k+1}} \sum_{u=v-1,v} \bar{\alpha}_k h_k(u,v) = \frac{1}{d_{k+1}} \sum_{u=v-1,v} \bar{\alpha}_k h_k(u,v).$$

Since

$$d_{k+1} = \sum_{0 \leq v \leq M_{k+1}} \sum_{u=v-1,v} \bar{\alpha}_k h_k(u,v),$$

we have the gradient using the chain rule

$$d'_{k+1} = \sum_{0 \leq v \leq M_{k+1}} \sum_{u=v-1,v} \left[\bar{\alpha}'_k h_k(u,v) + \bar{\alpha}_k h'_k(u,v) \right].$$

Using the chain rule again, the recursion of the gradient of the normalized forward variable is given by

$$\bar{\alpha}_{k+1}(v) = \frac{d'_{k+1}}{(d_{k+1})^2} \sum_{u=v-1,v} \bar{\alpha}_k h_k(u,v) + \frac{1}{d_{k+1}} \sum_{u=v-1,v} \left[\bar{\alpha}'_k h_k(u,v) + \bar{\alpha}_k h'_k(u,v) \right]. \tag{22}$$

The gradient of the log-likelihood is then

$$(\log \mathbb{L}_n)' = \sum_{k=1}^{n-1} \frac{d'_{k+1}}{d_{k+1}}.$$

R packages

All analyses, simulations, and computational implementations in this study were performed R version 4.4.2.[27]. Distribution fitting and goodness-of-fit diagnostics were performed using packages **fitdistrplus, goftest,** and **stats** [27–29]. Likelihood-based parameter estimation via differential evolution algorithm, gradient-based optimization, and automatic differentiation were carried out using **DEoptim**, **parallelly**, **base R**, and **TMB** [24, 26, 27, 30]. All figures were generated using **ggplot2** [31].
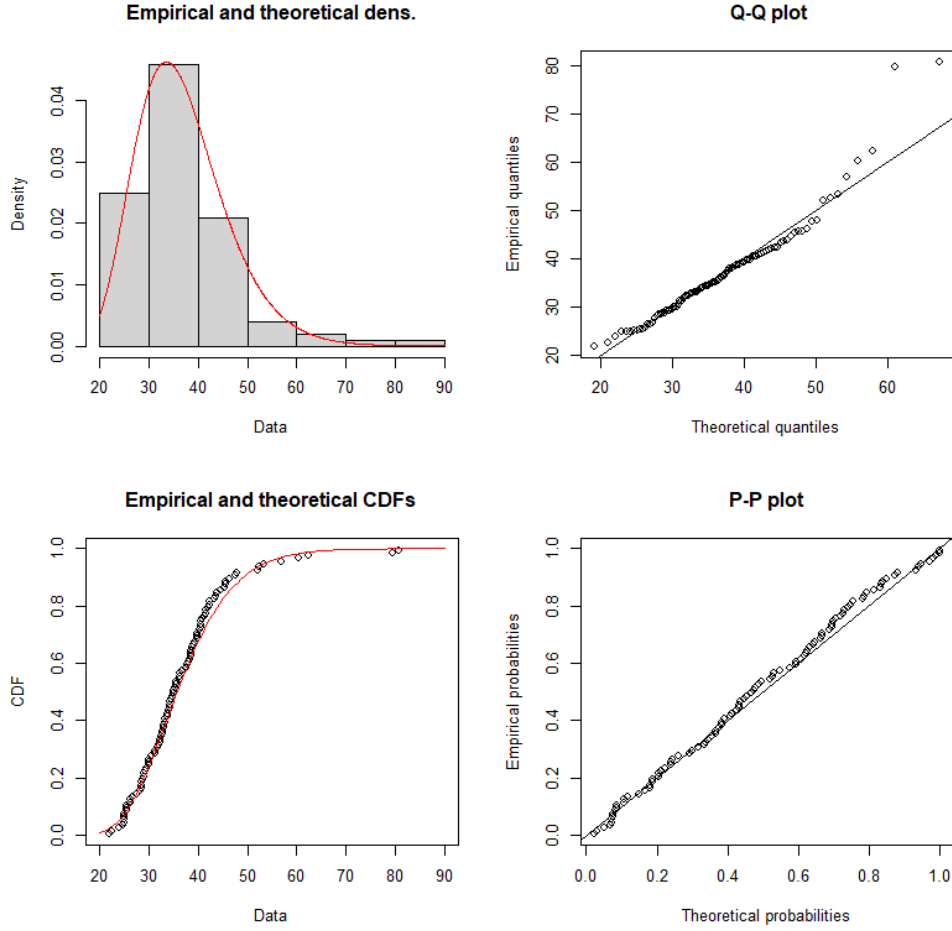
18

Fig. 5 Diagnostic plots assessing the fit of the inverse Guassian distribution to the stimulated stopping times. The fitted distribution is based on parameters estimated using maximum likelihood.

References

[1] Coletti, A.M., Singh, D., Kumar, S., Shafin, T.N., Briody, P.J., Babbitt, B.F., Pan, D., Norton, E.S., Brown, E.C., Kahle, K.T., *et al.*: Characterization of the ventricular-subventricular stem cell niche during human brain development. Development **145**(20), 170100 (2018)

[2] Bast, L., Calzolari, F., Strasser, M.K., Hasenauer, J., Theis, F.J., Ninkovic, J., Marr, C.: Increasing neural stem cell division asymmetry and quiescence are predicted to contribute to the age-related decline in neurogenesis. Cell Reports **25**(12), 3231–3240 (2018)

[3] Marciniak-Czochra, A., Stiehl, T., Wagner, W.: Modeling of replicative senescence in hematopoietic development. Aging (Albany NY) **1**(8), 723 (2009)

[4] Wilson, A., Laurenti, E., Oser, G., Wath, R.C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C.F., Eshkind, L., Bockamp, E., *et al.*: Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell **135**(6), 1118–1129 (2008)

[5] Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., Clarke, M.F.: Prospective identification of tumorigenic breast cancer cells. Proceedings of the National Academy of Sciences **100**(7), 3983–3988 (2003)

[6] Cammareri, P., Lombardo, Y., Francipane, M.G., Bonventre, S., Todaro, M., Stassi, G.: Isolation and culture of colon cancer stem cells. Methods in Cell Biology **86**, 311–324 (2008)

[7] Fioriti, D., Mischitelli, M., Monaco, F.D., Silverio, F.D., Petrangeli, E., Russo, G., Giordano, A., Pietropaolo, V.: Cancer stem cells in prostate adenocarcinoma: a target for new anticancer strategies. Journal of cellular physiology **216**(3), 571–575 (2008)

[8] Singh, S.K., Clarke, I.D., Terasaki, M., Bonn, V.E., Hawkins, C., Squire, J., Dirks, P.B.: Identification of a cancer stem cell in human brain tumors. Cancer Research **63**(18), 5821–5828 (2003)

[9] Beretta, E., Capasso, V., Morozova, N.: Mathematical modelling of cancer stem cells population behavior. Mathematical Modelling of Natural Phenomena **7**(1), 279–305 (2012)

[10] Weekes, S.L., Barker, B., Bober, S., Cisneros, K., Cline, J., Thompson, A., Hlatky, L., Hahnfeldt, P., Enderling, H.: A multicompartment mathematical model of cancer stem cell-driven tumor growth dynamics. Bulletin of Mathematical Biology **76**, 1762–1782 (2014)

[11] Chen, X., Wang, Y., Feng, T., Yi, M., Zhang, X., Zhou, D.: The overshoot and phenotypic equilibrium in characterizing cancer dynamics of reversible phenotypic plasticity. Journal of theoretical biology **390**, 40–49 (2016)

[12] Turner, C., Stinchcombe, A., Kohandel, M., Singh, S., Sivaloganathan, S.: Characterization of brain cancer stem cells: a mathematical approach. Cell Proliferation **42**(4), 529–540 (2009)

[13] Scanlon, V.M., Thompson, E.N., Lawton, B.R., Kochugaeva, M., Ta, K., Mayday, M.Y., Xavier-Ferrucio, J., Kang, E., Eskow, N.M., Lu, Y.-C., *et al.*: Multiparameter analysis of timelapse imaging reveals kinetics of megakaryocytic erythroid progenitor clonal expansion and differentiation. Scientific Reports **12**(1), 16218 (2022)

[14] Watson, H.W., Galton, F.: On the probability of the extinction of families. The Journal of the Anthropological Institute of Great Britain and Ireland **4**, 138–144 (1875)

[15] Harris, T.E.: Branching processes. The Annals of Mathematical Statistics, 474–494 (1948)

[16] Harris, T.E.: The Theory of Branching Processes. Springer, Berlin (1963)

[17] Asmussen, S., Hering, H.: Branching processes (1983)

[18] MacMillan, H.R., McConnell, M.J.: Seeing beyond the average cell: branching process models of cell proliferation, differentiation, and death during mouse brain development. Theory in Biosciences **130**(1), 31–43 (2011)

[19] Míguez, D.G.: A branching process to characterize the dynamics of stem cell differentiation. Scientific reports **5**(1), 13265 (2015)

[20] Folks, J.L., Chhikara, R.S.: The inverse Gaussian distribution and its statistical application—a review. Journal of the Royal Statistical Society Series B: Statistical Methodology **40**(3), 263–275 (1978)

[21] Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE **77**(2), 257–286 (2002)

[22] Storn, R., Price, K.: Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization **11**(4), 341–359 (1997)

[23] Ardia, D., Boudt, K., Carl, P., Mullen, K., Peterson, B.G.: Differential evolution with DEoptim: an application to non-convex portfolio optimization. The R Journal **3**(1), 27–34 (2011)

[24] Mullen, K.M., Ardia, D., Gil, D.L., Windover, D., Cline, J.: DEoptim: An R package for global optimization by differential evolution. Journal of Statistical Software **40**, 1–26 (2011)

[25] Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., Sibert, J.: AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software **27**(2), 233–249 (2012)

[26] Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., Bell, B.M.: TMB: automatic differentiation and Laplace approximation. Journal of statistical software **70**, 1–21 (2016)

[27] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2024). R Foundation for

Statistical Computing. https://www.R-project.org/

[28] Delignette-Muller, M.L., Dutang, C.: fitdistrplus: An R package for fitting distributions. Journal of Statistical Software **64**(4), 1–34 (2015) https://doi.org/10.18637/jss.v064.i04

[29] Faraway, J., Marsaglia, G., Marsaglia, J., Baddeley, A.: goftest: Classical Goodness-of-Fit Tests for Univariate Distributions. (2021). R package version 1.2-3. https://CRAN.R-project.org/package=goftest

[30] Bengtsson, H.: parallelly: Enhancing the 'parallel' Package. (2025). R package version 1.45. https://parallelly.futureverse.org

[31] Wickham, H.: ggplot2: Elegant Graphics for Data Analysis. Springer, New York (2016). https://ggplot2.tidyverse.org