Escape times for branching processes with random mutational fitness effects

Jasmine Foo[1], Kevin Leder[2], Junfeng Zhu[2]

1. School of Mathematics

2. Industrial and Systems Engineering,

University of Minnesota, Minneapolis, MN

June 2, 2014

Abstract

We consider a large declining population of cells under an external selection pressure, modeled as a subcritical branching process. This population has genetic variation introduced at a low rate which leads to the production of exponentially expanding mutant populations, enabling population escape from extinction. Here we consider two possible settings for the effects of the mutation: Case (I) a deterministic mutational fitness advance and Case (II) a random mutational fitness advance. We first establish a functional central limit theorem for the renormalized and sped up version of the mutant cell process. We establish that in Case (I) the limiting process is a trivial constant stochastic process, while in Case (II) the limit process is a continuous Gaussian process for which we identify the covariance kernel. Lastly we apply the functional central limit theorem and some other auxiliary results to establish a central limit theorem (in the large initial population limit) of the first time at which the mutant cell population dominates the population. We find that the limiting distribution is Gaussian in both Case (I) and (II), but a logarithmic correction is needed in the scaling for Case (II). This problem is motivated by the question of optimal timing for switching therapies to effectively control drug resistance in biomedical applications.

1 Introduction

Genetic variation often drives the process of population escape from extinction. For example, populations of bacteria or cancer cells declining under drug treatment can produce resistant variants capable of thriving under treatment, resulting in population rebound. Although new therapies are constantly being developed to target these drug-resistant mutants, one major question in the biomedical community today is: when should these second-line drugs be administered? Motivated by this question, here we consider a subcritical population of drug-sensitive cells in which a low rate of random genetic variation drives the production of a (possibly heterogeneous) population of resistant mutants. We are interested in studying the temporal dynamics of escape from extinction via this mechanism, and in particular here we obtain refined estimates of the stochastic time at which the resistant population first becomes dominant in the population. Characterization of this 'crossover' time, its variability, and how it depends on fundamental parameters of the drug profile and cell type, is useful in determining the optimal time to switch therapies and target different disease subpopulations. More generally, this work contributes to a growing literature aimed at developing theoretical tools for the design of dynamic treatment strategies that optimally utilize multiple drugs to control heterogeneous, evolving disease cell populations [9, 10, 4].

Random mutational fitness landscapes. We will consider a general setting where genetic variation can result in deterministic or random changes to the fitness of resistant cells. Under this setting genetic variation may produce a spectrum of effects on cellular fitness, resulting in a potentially highly heterogeneous population of resistant escape mutants. This type of intrinsic stochasticity in drug resistant populations has recently been a subject of intense biological interest and experimental investigations. For example, in a recent study experimentalists observed variability in inter-mitotic times in lung cancer cells with the T790M point mutation, which confers resistance to anti-cancer drugs erlotinib and gefitinib [12]. Another investigation revealed that within a clonal population of mycobacteria, there is significant heterogeneity among cells due to asymmetric cell division which renders them differentially resistant to several clinically important classes of antibiotics [1]. In light of these experimental developments, in this work we study the stochastic time of interest under cases where genetic variation produces both deterministic and random fitness effects in resistant cells, drawn from a mutational fitness landscape.

We build upon several previous related works. In the current investigation, we are interested in studying changes in the composition of the population which take place on a logarithmic time scale. Thus we utilize a time scaling considered in the works of Jagers, Sagitov, and Klebaner [8, 7], where the authors characterized process dynamics on the time scale of extinction of a subcritical branching process. In a previous work we established law of large numbers approximations of two escape times under this time scaling in the case of deterministic fitness effects [6]. In a joint work with Durrett, Mayberry and Michor [5], we also considered the impact of random mutational fitness effects on total population growth rate in expanding populations where multiple mutations are possible within the same cell. There it was shown that the addition of random fitness effects resulted in a polynomial time delay in the growth of the total population. Here we observe a consistent phenomenon, in that the addition of noise results in a decrease in the growth rate by a logarithmic term in the current time scale.

The main results in this paper are as follows. Theorem 1 establishes a law of large numbers approximation for the crossover time in the setting of random and deterministic fitness effects. Next, we prove a functional central limit theorem for the resistant cell population in Theorem 2. For the deterministic effect setting the limit is a degenerate stochastic process, whereas in the random fitness effect setting the limit is a continuous Gaussian process. Lastly, Theorem 3 establishes the weak limit of the crossover time in both settings.

The outline of the paper is as follows. In section 2 we introduce the model and define some quantities of interest. In section 3 the major results are stated, and the proof of the weak convergence result for the crossover time is provided. In section 4 the proofs of the remaining results are given. Throughout the paper we will use the following notation for the asymptotic behavior of positive functions.

$$\begin{aligned} f(t) \sim g(t) &\quad \text{if } f(t)/g(t) \to 1 \text{ as } t \to \infty \\ f(t) = O(g(t)) &\quad \text{if } f(t) \leq Cg(t) \text{ for all } t \\ f(t) = \Theta(g(t)) &\quad \text{if } cg(t) \leq |f(t)| \leq Cg(t) \text{ for all } t. \end{aligned}$$

In addition, C denotes a positive constant that may change throughout the paper.

2 Model and preliminaries

Consider a subcritical birth-death process Z_0 with birth rate r_0 death rate d_0 and net growth rate $\lambda_0 = r_0 - d_0 < 0$; this population represents the drug-sensitive cell population and is comprised of n cells at time $t = 0$. Assume that drug-resistant mutants are generated at time t at rate $Z_0(t)\mu n^{-\alpha}$ for $\alpha \in (0, 1)$, and that each of these mutations results in the creation of a supercritical birth-death process with random birth rate $d_0 + X$ and death rate d_0. Here, X is a possibly degenerate non-negative random variable with distribution G, and an independent copy of X is generated to determine the birth rate of each new mutant. Let us denote the total population of mutants as Z_1. Then, Z_1 is a supercritical branching process with immigration, which may be comprised of a spectrum of resistant types. We will consider two distinct types of distributions G:

1. *Case I (deterministic fitness effects)*:

$$G(dx) = \delta_{\lambda_1}(x)$$

2. *Case II (random fitness effects)*:

$$G(dx) = g(x)dx,$$

where $g(x)$ has support in $[0, \lambda_1]$ and is bounded, continuous and positive (at λ_1).

To clarify the model, we note that $Z_1(t)$ is a Markov process in Case I but not in Case II, since the population is heterogeneous.

For convenience we will also define $r = -\lambda_0$. We are interested in the asymptotic properties of the 'crossover time':

$$\xi_n = \inf\{t \geq 0 : Z_1(t) > Z_0(t)\},$$

i.e. the first time that the total population $Z_0 + Z_1$ is dominated by Z_1. Note that although Z_0 and Z_1 depend on n, we suppress the notation throughout the paper for the sake of notational simplicity.

Remark 1. *Note that we have used a mutation rate of $\mu n^{-\alpha}$ to model the behavior that each sensitive cell generates mutations at a very small rate (which may represent the rate of substitution error per base pair in DNA replication, for example). However, the initial population of sensitive cells is quite large in most biologically relevant settings. Therefore multiple simultaneous scalings are needed to allow for flexibility in the relationship between the large population size and the small mutation rate. This relationship can vary significantly between biological systems (e.g. variation between cell types, drug types, resistance mechanisms), and thus we are particularly interested in characterizing how the crossover time depends on the quantity α for biological applications.*

Remark 2. *Here we consider $\alpha \in (0, 1)$. The setting of $\alpha = 0$ is not biologically relevant since mutations are no longer rare and here the crossover time occurs on a time scale independent of n. The setting of $\alpha = 1$ is interesting and will be considered in future work. In this case only finitely many mutants are created and escape from extinction is no longer guaranteed.*

Remark 3. *An important class of distributions G that we do not explicitly consider are finite distributions, i.e., there is a finite set of points $x_1, \ldots, x_k$ and non-negative weights $p_1, \ldots, p_k$ summing to 1 such that for any $A \subset [0, \infty)$, $G(A) = \sum_{i=1}^k p_i 1_A(x_i)$. In this case the large time behavior of Z_1 will be largely determined by mutants with the fitness advance x_k. The dynamics of the crossover time in this case are similar to the results of Case I.*

In the remainder of this section we define some useful quantities that will be used throughout the paper. Let us define for $\phi_i(t) = \mathbb{E}Z_i(t)$ and $\psi_i(t) = \text{Var}Z_i(t)$, for $i = 0, 1$. One useful time scale in this problem is given by $t_n = \frac{1}{r}\log n$ which roughly approximates the time at which the Z_0 population dies out. On this time scale we have the following

$$\begin{aligned}
\phi_0(ut_n) &= n^{1-u} \\
\psi_0(ut_n) &= n^{1-u}\left(\frac{r_0 + d_0}{r}\right)\left(1 - n^{-u}\right).
\end{aligned}$$

For ease of the notation we will also introduce the following constants

$$\kappa_1 = \frac{2(d_0+\lambda_1)\mu}{\lambda_1(r+2\lambda_1)} \quad \text{and} \quad \kappa_2 = \frac{\mu r(\lambda_1+d_0)g(\lambda_1)}{\lambda_1(r+2\lambda_1)}.$$

For the type-1 population we have that

$$\begin{aligned} \phi_1(ut_n) &\sim m(n,u) \\ \psi_1(ut_n) &\sim \nu(n,u). \end{aligned}$$

where

$$m(n,u) = \begin{cases} \frac{\mu n^{1-\alpha+\lambda_1 u/r}}{\lambda_1+r}, & G(dx) = \delta_{\lambda_1}(x) \\ \frac{\mu r g(\lambda_1) n^{1-\alpha+\lambda_1 u/r}}{u(r+\lambda_1)\log n}, & G(dx) = g(x)dx, \end{cases}$$

and

$$\nu(n,u) = \begin{cases} \kappa_1 n^{1-\alpha+2\lambda_1 u/r}, & G(dx) = \delta_{\lambda_1}(x) \\ \frac{\kappa_2 n^{1-\alpha+2\lambda_1 u/r}}{u\log n}, & G(dx) = g(x)dx. \end{cases}$$

The unique positive root of $\phi_1(t) - \phi_0(t) = 0$ will be used as an approximation for ξ_n:

$$u_n \equiv \begin{cases} \frac{\log(1+n^\alpha(\lambda_1+r)/\mu)}{(\lambda_1+r)}, & G(dx) = \delta_{\lambda_1}(x) \\ u_*(n)t_n, & G(dx) = g(x)dx, \end{cases}.$$

Thus, $u_*(n)$ is the unique positive root to the equation: $\phi_1(ut_n) - \phi_0(ut_n) = 0$. We will see that the values of u_n/t_n for the two cases in fact converge to the same value as $n \to \infty$ in Proposition 1.

3 Results

We first establish the basic result of convergence in probability of the crossover time ξ_n to the estimate u_n under Cases I and II.

Theorem 1 (Convergence in probability for crossover time)**.** *For every $\varepsilon > 0$ we have that*

$$\lim_{n\to\infty} \mathbb{P}\left(|\xi_n - u_n| > \varepsilon\right) = 0.$$

Proof. See section 4.1. □

It should be noted that Theorem 1 is only used in the proof of Theorem 3.

Our overall goal is to identify a scaling $s_n \to \infty$ and a random variable χ such that $s_n(\xi_n - u_n) \Rightarrow \chi$. Later we will demonstrate that the appropriate value for this scaling is given by

$$s_n = \begin{cases} n^{(1-\alpha)/2}, & G(dx) = \delta_{\lambda_1}(x) \\ \sqrt{\frac{n^{1-\alpha}}{\log n}}, & G(dx) = g(x)dx. \end{cases} \tag{1}$$

In order to establish a weak convergence theorem for ξ_n, we need to first establish some additional results related to the fluctuations of Z_0 and Z_1.

Consider the fluctuations of the Z_1 process. To do this, define the centered and normalized process for $0 < u < 1$:

$$Y_n(u) = \frac{Z_1(ut_n) - \phi_1(ut_n)}{\sqrt{\nu(n,u)}}, \tag{2}$$

and the limiting covariance function for $0 < u \leq v < 1$

$$C(u,v) = \begin{cases} 1, & G(dx) = \delta_{\lambda_1}(x) \\ \frac{2\sqrt{uv}}{u+v}, & G(dx) = g(x)dx. \end{cases} \tag{3}$$

The next main result establishes a functional central limit theorem for the normalized and centered paths of Z_1.

Theorem 2 (FCLT for Z_1)**.** *For any $0 < a < b \leq 1$, $Y_n \Rightarrow Y$ in the standard Skorokhod topology on [a,b], where Y is a continuous centered Gaussian process on $[a,b]$ with covariance function $C(u,v)$, and $Y(a) \sim N(0,1)$.*

Proof. See section 4.2. □

Note that due to the large influx of resistance mutations at time $u = 0$ (in rescaled time), we were not able to establish tightness in the standard Skorokhod topology of the processes Y_n on time intervals including 0. Since we are primarily interested in 'crossover times' which occur at positive values of u, the behavior of the limit process at $u = 0$ is not important for this work; however we conjecture that there is a jump with probability one in the limit process at $u = 0$.

Our method for studying the fluctuations in ξ_n about u_n is to study the scaled maximum fluctuations of Z_0 and Z_1 about their means, and compare these quantities with the scaled maximum difference between ϕ_0 and ϕ_1. Our means of studying the difference in ϕ_0 and ϕ_1 is the expression

$$\sup_{u\in[a,u_n^-(y)]} \nu(n,u)^{-1/2} \left(\phi_1(ut_n) - \phi_0(ut_n)\right) \tag{4}$$

where $y > 0$, $u_n^-(y) = (u_n - y/s_n)/t_n$, and $0 < a < u_n^-(y)$ for n sufficiently large. We also define $u_n^+(y) = (u_n + y/s_n)/t_n$, but to avoid repetition we will generally focus our analysis on the supremum over $[a, u_n^-(y)]$, since the analysis of the supremum over $[a, u_n^+(y)]$ is nearly identical. In particular we have the following result.

Proposition 1 (Difference in means of Z_0 and Z_1)**.** *As $n \to \infty$,*

$$\sup_{u\in[a,u_n^-(y)]} \nu(n,u)^{-1/2} \left(\phi_1(ut_n) - \phi_0(ut_n)\right) \to \begin{cases} -\frac{y\mu}{\sqrt{\kappa_1}}, & G(dx) = \delta_{\lambda_1}(x) \\ -\mu y g(\lambda_1)\sqrt{\frac{r(\lambda_1+r)}{\alpha\kappa_2}}, & G(dx) = g(x)dx. \end{cases}$$

In addition we have that $u_*(n) \to \alpha r/(\lambda_1 + r)$.

Remark 4. *In the proof of Theorem 3 we need the observation that in Case II* $v(n,u)^{-1/2}|\phi_1(ut_n)-\phi_0(ut_n)|$ *grows rapidly as* u *moves away from* $u_*(n)$. *In particular following the proof of Proposition 1 (and in particular using display* (48)*), if* $\varepsilon_n = 1/\log n$ *it is possible to establish that*

$$\sup_{u\in[a,u_n^-(y)-\varepsilon_n]} \frac{\phi_1(ut_n)-\phi_0(ut_n)}{\sqrt{v(n,u)}} \tag{5}$$
$$= -r\mu e^{y(\lambda_1+r)/s_n}\sqrt{\frac{(u_n^-(y)-\varepsilon_n)n^{1-\alpha}}{\kappa_2\log n}}\int_0^{1+y/s_n}\frac{e^{-x(\lambda_1+r)}h_n\left(u_*(n)-x/t_n\right)}{u_*(n)-x/t_n}dx$$

where h_n *is bounded above 0 and defined in* (47). *Although* Y *will vary with* u *in Case II, this property will ensure that we only need to be concerned with its distribution at the limit of the crossover time, i.e.,* $u=(\alpha r)/(\lambda_1+r)$.

If time is sufficiently removed from the origin (on the logarithmic time scale), then we can safely ignore the fluctuations in the Z_0 population. In particular we have the following result

Proposition 2 (Fluctuations of Z_0)**.** *For* $a > (\frac{\alpha r}{\lambda_1+r})(\frac{\lambda_1+2r}{2(\lambda_1+r)})$ *we have that as* $n\to\infty$

$$\sup_{u\in[a,u_n^-(y)]} \nu(n,u)^{-1/2}(\phi_0(ut_n)-Z_0(ut_n))\to 0$$

in probability.

The proofs of the two propositions above are provided in sections 4.3 and 4.4. Using these results we can next establish the main result regarding the weak convergence limit of ξ_n.

Theorem 3 (Weak Convergence of ξ_n)**.** *If*

$$\sigma^2 = \begin{cases} \frac{2(d_0+\lambda_1)}{\lambda_1(r+2\lambda_1)\mu}, & G(dx)=\delta_{\lambda_1}(x) \\ \frac{\alpha(\lambda_1+d_0)}{\lambda_1\mu g(\lambda_1)(\lambda_1+r)(2\lambda_1+r)}, & G(dx)=g(x)dx \end{cases}$$

and $\chi\sim N(0,\sigma^2)$, *then as* $n\to\infty$,

$$s_n(\xi_n-u_n)\Rightarrow\chi$$

Proof. We will do this by studying the limits of the following probabilities

$$\lim_{n\to\infty}\mathbb{P}\left(s_n(\xi_n-u_n)<-y\right),$$

and

$$\lim_{n\to\infty}\mathbb{P}\left(s_n(\xi_n-u_n)>y\right),$$

for $y > 0$.

Now consider

$$\begin{aligned}\mathbb{P}\left(\xi_n < u_n - y/s_n\right) &= \mathbb{P}\left(\frac{\xi_n}{t_n} < u_n^-(y)\right) \\ &= \mathbb{P}\left(\sup_{u \le u_n^-(y)} \left(Z_1(ut_n) - Z_0(ut_n)\right) > 0\right), \end{aligned} \tag{6}$$

and similarly

$$\mathbb{P}\left(\xi_n > u_n + y/s_n\right) = \mathbb{P}\left(\sup_{u \le u_n^+(y)} \left(Z_1(ut_n) - Z_0(ut_n)\right) < 0\right).$$

First for $0 < a < u_n^-(y)$ we have the following inequalities

$$\begin{aligned}&\mathbb{P}\left(\sup_{u\in[a,u_n^-(y)]} \left(Z_1(ut_n) - Z_0(ut_n)\right) > 0\right) \le \mathbb{P}\left(\sup_{u\in[0,u_n^-(y)]} \left(Z_1(ut_n) - Z_0(ut_n)\right) > 0\right) \\ &\le \mathbb{P}\left(\sup_{u\in[0,a]} \left(Z_1(ut_n) - Z_0(ut_n)\right) > 0\right) + \mathbb{P}\left(\sup_{u\in[a,u_n^-(y)]} \left(Z_1(ut_n) - Z_0(ut_n)\right) > 0\right),\end{aligned}$$

and

$$\begin{aligned}&\mathbb{P}\left(\sup_{u\in[a,u_n^+(y)]} \left(Z_1(ut_n) - Z_0(ut_n)\right) < 0\right) - \mathbb{P}\left(\sup_{u\in[0,a]} \left(Z_1(ut_n) - Z_0(ut_n)\right) > 0\right) \\ &\le \mathbb{P}\left(\sup_{u\in[0,u_n^+(y)]} \left(Z_1(ut_n) - Z_0(ut_n)\right) < 0\right) \le \mathbb{P}\left(\sup_{u\in[a,u_n^+(y)]} \left(Z_1(ut_n) - Z_0(ut_n)\right) < 0\right).\end{aligned}$$

From Theorem 1 we know that $P(\sup_{u\in[0,a]}(Z_1(ut_n) - Z_0(ut_n)) > 0) \to 0$ as $n \to \infty$, and therefore it suffices to study the supremum over $u \in [a, u_n^-(y)]$ or $u \in [a, u_n^+(y)]$.

From now on we restrict ourselves to the analysis of the lower deviations, (6). The analysis of the upper deviations is nearly identical, and is thus omitted. First consider the bounds

$$\begin{aligned}&\mathbb{P}\left(Z_1(u_n^-(y)t_n) - Z_0(u_n^-(y)t_n) > 0\right) \le \mathbb{P}\left(\sup_{u\in[a,u_n^-(y)]} \left(Z_1(ut_n) - Z_0(ut_n)\right) > 0\right) \\ &\le \mathbb{P}\left(A_1(n,y) + A_2(n,y) + A_3(n,y) > 0\right),\end{aligned} \tag{7}$$

where

$$
\begin{aligned}
A_1(n,y) &= \sup_{u\in[a,u_n^-(y)]} \nu(n,u)^{-1/2}\left(Z_1(ut_n)-\phi_1(ut_n)\right)\\
A_2(n,y) &= \sup_{u\in[a,u_n^-(y)]} \nu(n,u)^{-1/2}\left(\phi_1(ut_n)-\phi_0(ut_n)\right)\\
A_3(n,y) &= \sup_{u\in[a,u_n^-(y)]} \nu(n,u)^{-1/2}\left(\phi_0(ut_n)-Z_0(ut_n)\right).
\end{aligned}
$$

For Case I, we can apply Theorem 2, Proposition 1, and Proposition 2 to see that

$$
\begin{aligned}
&\frac{Z_1\left(u_n^-(y)\right)-Z_0\left(u_n^-(y)\right)}{\nu(n,u)^{1/2}} \Rightarrow V - y\mu/\sqrt{\kappa_1}\\
&A_1(n,y)+A_2(n,y)+A_3(n,y) \Rightarrow V - y\mu/\sqrt{\kappa_1},
\end{aligned}
$$

where $V \sim N(0,1)$. Therefore the upper and lower bounds in (7) match and we see that

$$
\mathbb{P}\left(\sup_{u\in[a,u_n^-(y)]}\left(Z_1(ut_n)-Z_0(ut_n)\right)>0\right)\to \mathbb{P}\left(V>\frac{y\mu}{\sqrt{\kappa_1}}\right)
$$

which establishes the limit theorem. Note that this argument works in Case I since the limiting process, Y, is a single random variable and the distribution of the supremum is still Gaussian.

Case II requires a little more work. Here the upper and lower bounds from (7) no longer match since, $\sup_{u\in[a,b]} Y(u)$ and $Y(b)$ are no longer guaranteed to have the same distribution (since in Case II the covariance kernel is not identically 1). In the following we will evaluate the lower bound from (7) but obtain an improved matching upper bound. For the lower bound, we obtain once again from Theorem 2, Proposition 1, and Proposition 2 that

$$
\mathbb{P}\left(Z_1(u_n^-(y)t_n)-Z_0(u_n^-(y)t_n)>0\right)\to \mathbb{P}\left(V>y\mu g(\lambda_1)\sqrt{\frac{r(\lambda_1+r)}{\alpha\kappa_2}}\right)
$$

To obtain a matching upper bound, define $\varepsilon_n = 1/\log n$, the random variables

$$
\begin{aligned}
A_1(y,\varepsilon_n;n) &= \sup_{u\in[u_n^-(y)-\varepsilon_n,u_n^-(y)]} \nu(n,u)^{-1/2}\left(Z_1(ut_n)-\phi_1(ut_n)\right)\\
A_2(y,\varepsilon_n;n) &= \sup_{u\in[u_n^-(y)-\varepsilon_n,u_n^-(y)]} \nu(n,u)^{-1/2}\left(\phi_1(ut_n)-\phi_0(ut_n)\right)\\
A_3(y,\varepsilon_n;n) &= \sup_{u\in[u_n^-(y)-\varepsilon_n,u_n^-(y)]} \nu(n,u)^{-1/2}\left(\phi_0(ut_n)-Z_0(ut_n)\right),
\end{aligned}
$$

and the corresponding supremum over the remainder of the set $[a, u_n^-(y)]$,

$$A_1^c(y, \varepsilon_n; n) = \sup_{u \in [a, u_n^-(y) - \varepsilon_n]} \nu(n,u)^{-1/2} \left(Z_1(ut_n) - \phi_1(ut_n)\right)$$
$$A_2^c(y, \varepsilon_n; n) = \sup_{u \in [a, u_n^-(y) - \varepsilon_n]} \nu(n,u)^{-1/2} \left(\phi_1(ut_n) - \phi_0(ut_n)\right)$$
$$A_3^c(y, \varepsilon_n; n) = \sup_{u \in [a, u_n^-(y) - \varepsilon_n]} \nu(n,u)^{-1/2} \left(\phi_0(ut_n) - Z_0(ut_n)\right).$$

Then

$$\mathbb{P}\left(\sup_{u \in [a, u_n^-(y)]} \left(Z_1(ut_n) - Z_0(ut_n)\right) > 0 \right)$$
$$\leq \mathbb{P}\left(\sum_{i=1}^{3} A_i^c(y, \varepsilon_n; n) > 0 \right) + \mathbb{P}\left(\sum_{i=1}^{3} A_i(y, \varepsilon_n; n) > 0 \right).$$

From the analysis of $A_3(y; n)$ we know that $A_3(y, \varepsilon_n; n)$ and $A_3^c(y, \varepsilon_n; n)$ both converge to 0 in probability. Furthermore following the analysis of $A_2(y; n)$ we can see that

$$\lim_{n \to \infty} A_2(y, \varepsilon_n; n) = -y\mu g(\lambda_1) \sqrt{\frac{r(\lambda_1 + r)}{\alpha \kappa_2}}.$$

Similarly, based on Remark 4 and display (5) we see that there is a positive constant c such that

$$A_2^c(y, \varepsilon_n; n) \leq -c\sqrt{\frac{n^{1-\alpha}}{\log n}}. \tag{8}$$

From the tightness of the sequence of processes Y_n we also have the stochastic boundedness property

$$\lim_{K \to \infty} \sup_n \mathbb{P}\left(\sup_{u \in [a, u_n^-(y) - \varepsilon_n]} Y_n(u) > K \right) = 0,$$

which combined with the result (8) and the asymptotic negligibility of A_3^c gives that

$$\lim_{n \to \infty} \mathbb{P}\left(\sum_{i=1}^{3} A_i^c(y, \varepsilon_n; n) > 0 \right) = 0.$$

It then remains to study $A_1(y, \varepsilon_n; n)$, in particular to achieve a tight upper bound we need to establish that

$$A_1(y, \varepsilon_n; n) \Rightarrow Y(\alpha r/(\lambda_1 + r)). \tag{9}$$

First observe that the limit process Y is continuous with probability 1 on $[a, b] \subset (0, 1]$, due to the smoothness of the covariance function $C(\cdot, \cdot)$ on $[a, b] \times [a, b]$. From (45) we observe that $u_n^-(y) \to \alpha r/(\lambda_1 + r)$ and since the limit process is continuous we then have that as $n \to \infty$

$$Y_n(u_n^-(y)) \Rightarrow Y(\alpha r/(\lambda_1 + r)).$$

Next for

$$\delta \geq \frac{3r \log \log n}{2(\lambda_1 + r) \log n} + \frac{1}{s_n t_n} + \varepsilon_n,$$

observe that due to (45) we have that

$$[u_n^-(y) - \varepsilon_n, u_n^-(y)] \subset [\alpha r/(\lambda_1 + r) - \delta, \alpha r/(\lambda_1 + r) + \delta].$$

Therefore we have the inequalities for $x \in \mathbb{R}$

$$\mathbb{P}\left(Y_n(u_n^-(y)) > x\right) \leq \mathbb{P}\left(\sup_{u \in [u_n^-(y) - \varepsilon_n, u_n^-(y)]} Y_n(u) > x\right) \tag{10}$$
$$\leq \mathbb{P}\left(\sup_{u \in [\frac{\alpha r}{\lambda_1 + r} - \delta, \frac{\alpha r}{\lambda_1 + r} + \delta]} Y_n(u) > x\right).$$

Since the process Y is continuous we have that

$$\lim_{\delta \to 0} \sup_{u \in [\frac{\alpha r}{\lambda_1 + r} - \delta, \frac{\alpha r}{\lambda_1 + r} + \delta]} Y(u) = Y\left(\alpha r/(\lambda_1 + r)\right),$$

with probability 1, and therefore

$$\lim_{\delta \to 0} \sup_x \left| \mathbb{P}\left(\sup_{u \in [\frac{\alpha r}{\lambda_1 + r} - \delta, \frac{\alpha r}{\lambda_1 + r} + \delta]} Y(u) > x\right) - \mathbb{P}\left(Y\left(\alpha r/(\lambda_1 + r)\right) > x\right) \right| = 0. \tag{11}$$

If we send $n \to \infty$ in (10) we can apply Theorem 2 to the leftmost and rightmost terms of (10) to get

$$\mathbb{P}\left(Y\left(\frac{\alpha r}{\lambda_1 + r}\right) > x\right) \leq \liminf_{n \to \infty} \mathbb{P}\left(\sup_{u \in [u_n^-(y) - \varepsilon_n, u_n^-(y)]} Y_n(u) > x\right)$$
$$\leq \limsup_{n \to \infty} \mathbb{P}\left(\sup_{u \in [u_n^-(y) - \varepsilon_n, u_n^-(y)]} Y_n(u) > x\right) \leq \mathbb{P}\left(\sup_{u \in [\frac{\alpha r}{\lambda_1 + r} - \delta, \frac{\alpha r}{\lambda_1 + r} + \delta]} Y(u) > x\right),$$

we then get the desired result by sending $\delta \to 0$ and applying (11). □

4 Proofs

4.1 Proof of Theorem 1

Proof. For $\varepsilon > 0$ define

$$\hat{u}_n^-(\varepsilon) = \frac{u_n - \varepsilon}{t_n} \quad \text{and} \quad \hat{u}_n^+(\varepsilon) = \frac{u_n + \varepsilon}{t_n},$$

and note that

$$\begin{aligned}
\mathbb{P}\left(\xi_n < u_n - \varepsilon\right) &= \mathbb{P}\left(\frac{\xi_n}{t_n} < \hat{u}_n^-(\varepsilon)\right) \\
&= \mathbb{P}\left(\sup_{u \le \hat{u}_n^-(\varepsilon)} \left(Z_1(ut_n) - Z_0(ut_n)\right) > 0\right) \\
&\le \mathbb{P}\left(\hat{A}_1(n,\varepsilon) + \hat{A}_2(n,\varepsilon) + \hat{A}_3(n,\varepsilon) > 0\right),
\end{aligned}$$

where

$$\begin{aligned}
\hat{A}_1(n,\varepsilon) &= \sup_{u \le \hat{u}_n^-(\varepsilon)} n^{u-1}\left(Z_1(ut_n) - \phi_1(ut_n)\right) \\
\hat{A}_2(n,\varepsilon) &= \sup_{u \le \hat{u}_n^-(\varepsilon)} n^{u-1}\left(\phi_1(ut_n) - \phi_0(ut_n)\right) \\
\hat{A}_3(n,\varepsilon) &= \sup_{u \le \hat{u}_n^-(\varepsilon)} n^{u-1}\left(\phi_0(ut_n) - Z_0(ut_n)\right).
\end{aligned}$$

Lemma 2 of [6] shows that $\hat{A}_3(n,\varepsilon) \to 0$ in probability. For G of Case I we have that $\hat{A}_1(n,\varepsilon) \to 0$ via a simplification of the argument in Theorem 5 of [6]. In Case II consider the bound

$$\begin{aligned}
\hat{A}_1(n,\varepsilon) &\le \sup_{u \le \hat{u}_n^-(\varepsilon)} n^{u(1+\lambda_1/r)-\alpha} \sup_{u \le \hat{u}_n^-(\varepsilon)} n^{\alpha-\lambda_1 u/r-1}|Z_1(ut_n) - \phi_1(ut_n)| \\
&\le n^{u_*(n)(\lambda_1+r)/r-\alpha} \sup_{u \le \hat{u}_n^-(\varepsilon)} n^{\alpha-\lambda_1 u/r-1}|Z_1(ut_n) - \phi_1(ut_n)|.
\end{aligned}$$

From Proposition 1 we know that $u_*(n)(\lambda_1 + r)/r - \alpha \to 0$, and thus for any $\eta < 0$ we have that

$$n^\eta n^{u_*(n)(\lambda_1+r)/r-\alpha} \to 0 \tag{12}$$

as $n \to \infty$. Next observe that

$$\begin{aligned}
&\sup_{u \le \hat{u}_n^-(\varepsilon)} n^{\alpha-\lambda_1 u/r-1}|Z_1(ut_n) - \phi_1(ut_n)| \\
&= \sqrt{\frac{n^{\alpha-1}}{\log n}} \sup_{u \le \hat{u}_n^-(\varepsilon)} \sqrt{\frac{\log n}{n^{1-\alpha+2\lambda_1 u/r}}}|Z_1(ut_n) - \phi_1(ut_n)|.
\end{aligned}$$

Recall the centered, rescaled process Y_n from equation (2). From Theorem 2 we know that the sequence Y_n is relatively compact. Therefore Theorem 13.2 of [3] implies that for $0 < a < b < 1$

$$\sup_n \mathbb{P}\left(\sup_{u\in[a,b]} \sqrt{\frac{\log n}{n^{1-\alpha+2\lambda_1 u/r}}}|Z_1(ut_n) - \phi_1(ut_n)| > K\right) \to 0$$

as $K \to \infty$. Then combining the previous display with (12), we get that

$$\hat{A}_1(n,\varepsilon) \le n^{u_*(n)(\lambda_1+r)/r-\alpha}\sqrt{\frac{n^{\alpha-1}}{\log n}} \sup_{u\le \hat{u}_n^-(\varepsilon)} \sqrt{\frac{\log n}{n^{1-\alpha+2\lambda_1 u/r}}}|Z_1(ut_n) - \phi_1(ut_n)|$$

converges to 0 in probability.

It now remains to establish that $\hat{A}_2(n,\varepsilon)$ is a negative number bounded away from zero. First note that via monotonicity and the result $\phi_0(ut_n) = n^{1-u}$, $\hat{A}_2(n,\varepsilon)$ can be simplified to

$$\begin{aligned}
\hat{A}_2(n,\varepsilon) &= n^{\hat{u}_n^-(\varepsilon)-1}\left(\phi_1\left(\hat{u}_n^-(\varepsilon)t_n\right) - \phi_0\left(\hat{u}_n^-(\varepsilon)t_n\right)\right)\\
&= n^{\hat{u}_n^-(\varepsilon)-1}\phi_1\left(\hat{u}_n^-(\varepsilon)t_n\right) - 1.
\end{aligned}$$

The desired result will then follow by establishing that $n^{\hat{u}_n^-(\varepsilon)-1}\phi_1\left(\hat{u}_n^-(\varepsilon)t_n\right) < 1$. Thus consider

$$\begin{aligned}
n^{\hat{u}_n^-(\varepsilon)-1}\phi_1\left(\hat{u}_n^-(\varepsilon)t_n\right) &= n^{\hat{u}_n^-(\varepsilon)-1}\left(\phi_1(u_*(n)t_n) + \phi_1\left(\hat{u}_n^-(\varepsilon)t_n\right) - \phi_1(u_*(n)t_n)\right)\\
&= n^{u_*(n)-1}\phi_0(u_*(n)t_n)n^{\hat{u}_n^-(\varepsilon)-u_*(n)}\\
&\quad + n^{\hat{u}_n^-(\varepsilon)-1}\left(\phi_1\left(\hat{u}_n^-(\varepsilon)t_n\right) - \phi_1(u_*(n)t_n)\right)\\
&= e^{-\varepsilon r} + n^{\hat{u}_n^-(\varepsilon)-1}\left(\phi_1\left(\hat{u}_n^-(\varepsilon)t_n\right) - \phi_1(u_*(n)t_n)\right)\\
&\le e^{-\varepsilon r},
\end{aligned}$$

where the second equality is from the definition of $u_*(n)$, the third equality from the formula $\phi_0(ut_n) = n^{1-u}$ and the definition of $\hat{u}_n^-(\varepsilon)$, and the final inequality is due to the monotone increasing property of ϕ_1.

It now remains to study

$$\begin{aligned}
\mathbb{P}\left(\xi_n > u_n + \varepsilon\right) &= \mathbb{P}\left(\sup_{u\le \hat{u}_n^+(\varepsilon)} \left(Z_1(ut_n) - Z_0(ut_n)\right) < 0\right)\\
&\le \mathbb{P}\left(Z_1\left(\hat{u}_n^+(\varepsilon)t_n\right) - Z_0\left(\hat{u}_n^+(\varepsilon)t_n\right) < 0\right)\\
&= \mathbb{P}\left(B_1(n,\varepsilon) + B_2(n,\varepsilon) + B_3(n,\varepsilon) < 0\right),
\end{aligned}$$

where

$$\begin{aligned}
B_1(n,\varepsilon) &= n^{\hat{u}_n^+(\varepsilon)-1}\left(Z_1(\hat{u}_n^+(\varepsilon)t_n) - \phi_1(\hat{u}_n^+(\varepsilon)t_n)\right)\\
B_2(n,\varepsilon) &= n^{\hat{u}_n^+(\varepsilon)-1}\left(\phi_1(\hat{u}_n^+(\varepsilon)t_n) - \phi_0(\hat{u}_n^+(\varepsilon)t_n)\right)\\
B_3(n,\varepsilon) &= n^{\hat{u}_n^+(\varepsilon)-1}\left(\phi_0(\hat{u}_n^+(\varepsilon)t_n) - Z_0(\hat{u}_n^+(\varepsilon)t_n)\right).
\end{aligned}$$

Again using Lemma 2 and Theorem 5 from [6] we can show that as $n \to \infty$ $B_3(n,\varepsilon) \to 0$ and $B_1(n,\varepsilon) \to 0$. Also similar to our analysis of $\hat{A}_2(n,\varepsilon)$ we can show that for sufficiently large n, $B_2(n,\varepsilon) > 0$. The result then follows. □

4.2 Proof of Theorem 2

We first establish the weak convergence of the finite dimensional distributions of Y_n.

Lemma 1 (Limiting correlation)**.** *For any $k \in \mathbb{N}$ and $a \le u_1 < \cdots < u_k \le b$,*

$$(Y_n(u_1), \ldots, Y_n(u_k)) \Rightarrow (Y(u_1), \ldots, Y(u_k)),$$

where Y is a Gaussian process on $[a,b]$ with covariance function $C(u,v)$ given in (3) *and $Y(a) \sim N(0,1)$.*

In Case I we have the following result

Lemma 2 (Tightness Case I)**.** *In Case I, for any $0 < a < b < 1$ the sequence of processes $\{Y_n\}$ is tight in the standard Skorokhod topology for càdlàg functions on $[a,b]$.*

The previous two lemmas then establish Theorem 2 in Case I. To establish the result for Case II, it remains to establish tightness of the processes $\{Y_n\}$. For convenience we will work with the process

$$\tilde{Y}_n(u) = \left(n^{\alpha-1}t_n\right)^{1/2}\left(e^{-\lambda_1 u t_n} Z_1(ut_n) - \mu n^{1-\alpha}\int_0^{\lambda_1} e^{ut_n(x-\lambda_1)}\int_0^{ut_n} e^{-s(r+x)}g(x)dsdx\right), \tag{13}$$

which is a constant multiple of $Y_n(u)/\sqrt{u}$; thus establishing tightness for the processes $\{\tilde{Y}_n\}$ is sufficient.

Due to the non-Markovian nature of $\tilde{Y}_n$, in order to study tightness it is necessary to introduce an approximating process. Specifically for $\{\triangle_n\}$, a sequence of positive numbers decreasing to zero, define $\lambda_1(\triangle_n) = \text{ceil}(\lambda_1/\triangle_n)$, $x_j = j\triangle_n$ and $g_j = \int_{x_{j-1}}^{x_j} g(x)dx$ for $j \in \{0,1,\ldots,\lambda_1(\triangle_n)\}$. In addition define $\hat{Z}_{1,j}$ to be a branching process with birth rate $d_0 + x_j$, death rate d_0, initial size 0, and immigration rate at time s given by $\mu n^{-\alpha} Z_0(s) g_j$. With these definitions in place we now define an approximating process

$$J_n(u) = \left(n^{\alpha-1}t_n\right)^{1/2}\sum_{j=1}^{\lambda_1(\triangle_n)} e^{ut_n(x_j-\lambda_1)}\left(\hat{Z}_{1,j}(ut_n)e^{-ut_n x_j} - \frac{\mu g_j}{n^\alpha}\int_0^{ut_n} e^{-sx_j} Z_0(s)ds\right). \tag{14}$$

We will create a coupling between J_n and $\tilde{Y}_n$ to establish the following Lemma.

Lemma 3 (Approximating Process for Case II)**.** *Fix* $0 < a < b < 1$*. If* $\triangle_n = o\left(n^{\alpha-1}e^{-\lambda_1 b t_n}/t_n\right)$ *then for any* $\varepsilon > 0$ *we have that*

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{u\in[a,b]} |\tilde{Y}_n(u) - J_n(u)| > \varepsilon\right) = 0.$$

With the previous lemma we can work with the simpler processes $\{J_n\}$ and establish that

Lemma 4 (Tightness Case II)**.** *If* $\triangle_n = O(n^{2(\alpha-1)/3}/t_n^{4/3})$ *then for any* $0 < a < b \leq 1$ *the sequence* $\{J_n\}$ *is tight in the standard Skorokhod topology for càdlàg functions on* $[a, b]$

We lastly need the following Lemma which is quite standard, but we could not find a proof in the literature so we provide one here.

Lemma 5. *Let* $B(t)$ *be Markovian continuous time branching process with offspring generating function* $f(s)$*,* $B(0) = 1$*, birth events occur at rate* a*, and define* $\lambda = a(f'(1) - 1)$*. If for non-negative integer* k*,* $f^{(k)}(1) < \infty$ *then we have that*

$$E[B(t)^k] = \begin{cases} \Theta(e^{\lambda k t}), & \lambda > 0 \\ \Theta(e^{\lambda t}), & \lambda < 0 \end{cases}$$

With this result we are now ready to complete the proof of the functional central limit theorem.

Proof of Theorem 2. As mentioned above, in Case I the proof is complete by combining Lemmas 1 and 2. For Case II it suffices to establish weak convergence of $\{\tilde{Y}_n\}$. From Lemma 1 we get finite dimensional distribution (FDD) convergence of $\{\tilde{Y}_n\}$. Then via Slutsky's theorem and Lemma 3 we have FDD convergence of $\{J_n\}$. We then get weak convergence of $\{J_n\}$ via Lemma 4, which of course implies weak convergence of $\{\tilde{Y}_n\}$ via Slutsky's theorem and Lemma 3. □

4.2.1 Proof of Lemma 1

Proof. We first identify the limit of the following correlation function

$$C_n(u, v) \doteq \frac{E\left[\left(Z_1(ut_n) - \phi_1(ut_n)\right)\left(Z_1(vt_n) - \phi_1(vt_n)\right)\right]}{\left(\nu(n, u)\nu(n, v)\right)^{1/2}}, \tag{15}$$

i.e., to identify $\lim_{n\to\infty} C_n(u, v) = C(u, v)$. We will first consider this in the case $G(dx) = \delta_{\lambda_1}(x)$, and in this setting observe that if we define $\mathcal{F}_t = \sigma\left((Z_0(s), Z_1(s)), s \leq t\right)$ then for $u \leq v$

$$E[Z_1(vt_n)|\mathcal{F}_{ut_n}] = e^{\lambda_1 t_n(v-u)} Z_1(ut_n) + \frac{e^{\lambda_1 v t_n}\mu}{n^\alpha}\int_{ut_n}^{vt_n} e^{-\lambda_1 s} E[Z_0(s)|\mathcal{F}_{ut_n}]ds.$$

By conditioning we have

$$E[Z_1(ut_n)Z_1(vt_n)] = e^{\lambda_1 t_n(v-u)}E[Z_1(ut_n)^2] + \frac{e^{\lambda_1 vt_n}\mu}{n^\alpha}\int_{ut_n}^{vt_n} e^{-\lambda_1 s}E[Z_0(s)Z_1(ut_n)]ds,$$

then applying the first part of Lemma 1 of [6]

$$\begin{aligned}E[Z_1(ut_n)Z_1(vt_n)] &= \frac{e^{\lambda_1 t_n(v+u)}\mu^2}{n^{2\alpha}}\int_0^{ut_n}\int_0^{ut_n} e^{-\lambda_1(s+y)}E[Z_0(s)Z_0(y)]dsdy \\ &+ \mu n^{1-\alpha}e^{\lambda_1 t_n(v-u)}\int_0^{ut_n} e^{-rs}E[\tilde{Z}_1(ut_n-s)^2]ds \\ &+ \frac{e^{\lambda_1 vt_n}\mu}{n^\alpha}\int_{ut_n}^{vt_n} e^{-\lambda_1 s}E[Z_0(s)Z_1(ut_n)]ds,\end{aligned}$$

and then applying the second part of Lemma 1 of [6] we have

$$\begin{aligned}E[Z_1(ut_n)Z_1(vt_n)] &= \frac{e^{\lambda_1 t_n(v+u)}\mu^2}{n^{2\alpha}}\int_0^{ut_n}\int_0^{ut_n} e^{-\lambda_1(s+y)}E[Z_0(s)Z_0(y)]dsdy \\ &+ \mu n^{1-\alpha}e^{\lambda_1 t_n(v-u)}\int_0^{ut_n} e^{-rs}E[\tilde{Z}_1(ut_n-s)^2]ds \\ &+ \frac{\mu^2 e^{\lambda_1 t_n(u+v)}}{n^{2\alpha}}\int_{ut_n}^{vt_n}\int_0^{ut_n} e^{-\lambda_1(y+s)}E[Z_0(y)Z_0(s)]dyds,\end{aligned}$$

where $\tilde{Z}_1$ is binary branching process with birth rate $d_0+\lambda_1$, death rate d_0, and $\tilde{Z}_1(0)=1$. Next observe that

$$\begin{aligned}\phi_1(ut_n)\phi_1(vt_n) &= \frac{\mu^2 e^{\lambda_1 t_n(u+v)}}{n^{2\alpha}}\int_0^{ut_n}\int_0^{ut_n} E[Z_0(s)]E[Z_0(y)]e^{-\lambda_1(s+y)}dsdy \\ &+ \frac{\mu^2 e^{\lambda_1 t_n(u+v)}}{n^{2\alpha}}\int_{ut_n}^{vt_n}\int_0^{ut_n} E[Z_0(s)]E[Z_0(y)]e^{-\lambda_1(s+y)}dyds,\end{aligned}$$

and therefore

$$\begin{aligned}&E\left[\left(Z_1(ut_n)-\phi_1(ut_n)\right)\left(Z_1(vt_n)-\phi_1(vt_n)\right)\right] \\ &= \frac{e^{\lambda_1 t_n(v+u)}\mu^2}{n^{2\alpha}}\int_0^{ut_n}\int_0^{ut_n} e^{-\lambda_1(s+y)}\mathrm{Cov}[Z_0(s)Z_0(y)]dsdy \\ &+ \frac{\mu^2 e^{\lambda_1 t_n(u+v)}}{n^{2\alpha}}\int_{ut_n}^{vt_n}\int_0^{ut_n} e^{-\lambda_1(y+s)}\mathrm{Cov}[Z_0(y)Z_0(s)]dyds \\ &+ \mu n^{1-\alpha}e^{\lambda_1 t_n(v-u)}\int_0^{ut_n} e^{-rs}E[\tilde{Z}_1(ut_n-s)^2]ds.\end{aligned}$$

Since $\mathrm{Cov}[Z_0(s)Z_0(y)] = O(n)$ and $E[\tilde{Z}_1(ut_n-s)^2] = O(e^{2\lambda_1(ut_n-s)})$ (see Lemma 5) we see that the first two terms in the previous expression are $O(n^{1-2\alpha}e^{\lambda_1 t_n(u+v)})$, while the latter expression is $O(n^{1-\alpha}e^{\lambda_1 t_n(u+v)})$. Therefore when analyzing the limit of $C_n(u,v)$ it follows that the only the final term will impact the limit. Thus it remains to analyze the limit below, which is a straightforward calculation

$$\lim_{n\to\infty} \frac{\mu n^{1-\alpha}e^{\lambda_1 t_n(v-u)}}{(\nu(n,u)\nu(n,v))^{1/2}} \int_0^{ut_n} e^{-rs}E[\tilde{Z}_1(ut_n-s)^2]ds = 1.$$

Thus we see that in the case of deterministic advances that the limiting correlation function is $C(u,v)=1$.

We now consider the case of random mutational advances, i.e., $G(dx)=g(x)dx$. Following a similar development as in the deterministic advance we calculate the covariance for $0<u\le v\le 1$. In particular, by discretizing the fitness space and time and using the independence of distinct cell lines we obtain:

$$\begin{aligned}
&\text{Cov}(Z_1(ut_n), Z_1(vt_n)) \\
&= 2\mu n^{1-\alpha}\int_0^{\lambda_1} \frac{g(x)(d_0+x)e^{xvt_n}}{x(r+2x)}\left[e^{xut_n} - \frac{(2d_0+x)(r+2x)}{2(d_0+x)(r+x)}\right]dx(1+o(1)).
\end{aligned}$$

We will show that

$$\frac{2\mu n^{1-\alpha}}{(\nu(n,u)\nu(n,v))^{1/2}}\int_0^{\lambda_1/2} \frac{g(x)(d_0+x)e^{xvt_n}}{x(r+2x)}\left[e^{xut_n} - \frac{(2d_0+x)(r+2x)}{2(d_0+x)(r+x)}\right]dx, \qquad (16)$$

goes to 0 as $n\to\infty$. First note that

$$e^{xut_n} - \frac{(2d_0+x)(r+2x)}{2(d_0+x)(r+x)} = e^{xut_n} - 1 - \frac{x(2d_0-r)}{2(d_0+x)(r+x)},$$

so by the mean value theorem

$$\frac{1}{x}\left|e^{xut_n} - \frac{(2d_0+x)(r+2x)}{2(d_0+x)(r+x)}\right| \le ut_n e^{xut_n} + \frac{|2d_0-r|}{2d_0 r} \le Cut_n e^{xut_n}$$

for a positive constant C. With this bound we can bound (16) by

$$\frac{2Cut_n\mu n^{1-\alpha}}{(\nu(n,u)\nu(n,v))^{1/2}}\int_0^{\lambda_1/2} \frac{g(x)(d_0+x)e^{xt_n(u+v)}}{x(r+2x)}dx = O\left(\frac{t_n n^{1-\alpha}e^{\lambda_1 t_n(u+v)/2}}{(\nu(n,u)\nu(n,v))^{1/2}}\right),$$

which goes to 0 as $n\to\infty$.

Thus it suffices to consider

$$\frac{2\mu n^{1-\alpha}}{(\nu(n,u)\nu(n,v))^{1/2}}\int_{\lambda_1/2}^{\lambda_1} \frac{g(x)(d_0+x)e^{xvt_n}}{x(r+2x)}\left[e^{xut_n} - \frac{(2d_0+x)(r+2x)}{2(d_0+x)(r+x)}\right]dx,$$

and observe that the second term in the integral clearly goes to 0. Using the change of variable $y = t_n(u+v)(\lambda_1 - x)$ and plugging in the definition of $\nu(n,u)$ we see that

$$\lim_{n\to\infty} \frac{2\mu n^{1-\alpha}}{(\nu(n,u)\nu(n,v))^{1/2}} \int_{\lambda_1/2}^{\lambda_1} \frac{g(x)(d_0+x)e^{xt_n(v+u)}}{x(r+2x)} dx = \frac{2\sqrt{uv}}{(u+v)}.$$

We now establish the asymptotic normality of the large n limit of $(Y_n(u), Y_n(v))$ for $0 < u \le v \le 1$. The proof for more than two time points will be identical thus we only consider the setting of two time points. By the Cramer-Wold device it suffices to study weak limit of

$$\theta_1 Y_n(u) + \theta_2 Y_n(v)$$

for arbitrary $(\theta_1, \theta_2) \in \mathbb{R}^2$. Recall that $Z_0(0) = n$, and use the label $Z_0(t) = \sum_{j=1}^n Z_0^{(j)}(t)$ where $Z_0^{(j)}$ are the un-mutated offspring of cell j from the original population. Further write $Z_1(t) = \sum_{j=1}^n Z_1^{(j)}(t)$, where $Z_1^{(j)}$ are the mutated offspring of cell j, in addition define $\phi_1^{(j)}(ut_n) = \phi_1(ut_n)/n$. If we then define

$$Y_n^{(j)}(u) = \frac{Z_1^{(j)}(ut_n) - \phi_1^{(j)}(ut_n)}{\nu(n,u)^{1/2}}$$

then it follows that

$$\theta_1 Y_n(u) + \theta_2 Y_n(v) = \sum_{j=1}^n \left(\theta_1 Y_n^{(j)}(u) + \theta_2 Y_n^{(j)}(v)\right).$$

Then

$$C(u,v) = \lim_{n\to\infty} n \mathrm{Cov}\left(Y_n^{(j)}(u), Y_n^{(j)}(v)\right),$$

whose value we calculated earlier within this proof. If we establish the following Lindeberg condition

$$\lim_{n\to\infty} n\mathbb{E}\left[\left(\theta_1 Y_n^{(j)}(u) + \theta_2 Y_n^{(j)}(v)\right)^2; |\theta_1 Y_n^{(j)}(u) + \theta_2 Y_n^{(j)}(v)| > \varepsilon\right] = 0, \tag{17}$$

then

$$\theta_1 Y_n(u) + \theta_2 Y_n(v) \Rightarrow Z(u,v;\theta_1,\theta_2),$$

where $Z(u,v;\theta_1,\theta_2) \sim N(0,\omega^2)$ with $\omega^2 = \theta_1^2 + \theta_2^2 + 2\theta_1\theta_2 C(u,v)$. Since the following analysis applies to arbitrary $(\theta_1,\theta_2) \in \mathbb{R}^2$ it follows that $(Y_n(u), Y_n(v))$ converges in distribution to a mean-zero Gaussian with covariance matrix given by

$$\Sigma = \begin{pmatrix} 1 & C(u,v) \\ C(u,v) & 1 \end{pmatrix}$$

Thus it remains to establish (17). By expanding the square applying the Cauchy-Schwarz inequality, and the inequality

$$1_{\{|\theta_1 Y_n^{(j)}(u) + \theta_2 Y_n^{(j)}(v)| > \varepsilon\}} \le 1_{\{|\theta_1 Y_n^{(j)}(u)| \ge \varepsilon/2\}} + 1_{\{|\theta_2 Y_n^{(j)}(v)| \ge \varepsilon/2\}}$$

we see that (17) will be implied by establishing

$$\lim_{n\to\infty} nE\left[Y_n^{(j)}(u)^2; |Y_n^{(j)}(v)| > \varepsilon\right] = 0 \tag{18}$$

for arbitrary $(u,v) \in (0,1)\times(0,1)$. We will only do this for the setting $G(dx) = g(x)dx$ since the deterministic advance case is a simpler version of the same arguments. Observe that in this setting

$$nY_n^{(j)}(u)^2 = \frac{u\log n}{\kappa_2 n^{2\lambda_1 u/r-\alpha}}\left(Z_1^{(j)}(ut_n) - \phi_1^{(j)}(ut_n)\right)^2.$$

Denote the number of mutations created by cell j by time ut_n by $N_j(ut_n)$ which is a Poisson process with intensity at time s given by $\mu n^{-\alpha} Z_0^{(j)}(s)$. Further denote the time of creation of the ith mutant by cell j by τ_{ij} and the net growth rate of this cell by X_{ij}. Lastly denote the Markovian branching process representing the descendants of the ith mutant of initial cell j by $B_i^{(j)}$, note that this population will have death rate d_0 and birth rate $d_0 + X_{ij}$. Based on the above notation we have

$$Z_1^{(j)}(ut_n) = \sum_{i=1}^{N_j(ut_n)} B_i^{(j)}(ut_n - \tau_{ij}). \tag{19}$$

Now choose $\eta \in (0, \lambda_1)$, and form the quantities

$$\begin{aligned}
Z_{1,\eta}^{(j)}(ut_n) &= \sum_{i=1}^{N_j(ut_n)} B_i^{(j)}(ut_n - \tau_{ij}) 1_{\{X_{ij}\le\eta\}} \\
\phi_{1,\eta}^{(j)}(ut_n) &= \frac{\mu}{n^\alpha}\int_0^\eta g(x)\int_0^{ut_n} e^{x(ut_n-s)} Z_0^{(j)}(s)dsdx \\
Y_{n,\eta}^{(j)}(u) &= \frac{Z_{1,\eta}^{(j)}(ut_n) - \phi_{1,\eta}^{(j)}(ut_n)}{\nu(n,u)^{1/2}}.
\end{aligned}$$

It is then immediate that $nE[Y_{n,\eta}^{(j)}(u)^2] \to 0$ since $\mathrm{Var}(Z_{1,\eta}^{(j)}(ut)) = O(e^{2\eta ut_n} n^{-\alpha}/\log n)$ and $\nu(n,u) = \Theta(e^{2\lambda_1 ut_n} n^{1-\alpha}/\log n)$. Thus we assume that without loss of generality that all mutational advances confer a fitness advance greater than η.

Note that for any Markov branching process B with mean growth rate $\lambda > 0$ and offspring generating function f with $f''(1) < \infty$ there exists a square integrable random variable W such that $E[(W - e^{-\lambda t}B(t))^2] \to 0$ as $t \to \infty$ (e.g., see Theorem 2 of I.6 in [2]). This of course implies that $E[W^2] = f''(1)/(f'(1) - 1)$. Further note that since $\sup_{t\ge0} e^{-2\lambda t}E[B(t)^2] < \infty$, we see that $e^{-\lambda t}B(t)$ is a uniformly integrable martingale, and in particular if $\mathcal{G}_t = \sigma(B(s), s \le t)$, then $E[W|\mathcal{G}_t] = B(t)e^{-\lambda t}$.

Therefore

$$E[(B(t)e^{-\lambda t}-W)^2] = E[W^2] - e^{-2\lambda t}E[B(t)^2] \tag{20}$$
$$= \frac{f''(1)}{f'(1)-1} - e^{-2\lambda t}\left(\frac{f''(1)}{f'(1)-1}e^{2\lambda t} - \frac{f''(1)-(f'(1)-1)}{f'(1)-1}e^{\lambda t}\right)$$
$$= \frac{f''(1)-(f'(1)-1)}{f'(1)-1}e^{-\lambda t} = \kappa(\lambda)e^{-\lambda t}.$$

Observe that $\sup_{x\in[\eta,\lambda_1]}\kappa(x) = K < \infty$. Thus for each process $B_i^{(j)}$ let W_{ij} be the L_2 limiting random variable. Then

$$\frac{u\log n}{\kappa_2 n^{2\lambda_1 u/r-\alpha}}\left(Z_1^{(j)}(ut_n) - \phi_1^{(j)}(ut_n)\right)^2$$
$$\leq \frac{2n^{\alpha}u\log n}{\kappa_2}\left(\sum_{i=1}^{N_j(ut_n)} e^{ut_n(X_{ij}-\lambda_1)}e^{-X_{ij}\tau_{ij}}W_{ij} - \phi_1^{(j)}(ut_n)n^{-\lambda_1 u/r}\right)^2$$
$$+ \frac{2n^{\alpha}u\log n}{\kappa_2}\left(\sum_{i=1}^{N_j(ut_n)} e^{ut_n(X_{ij}-\lambda_1)}e^{-X_{ij}\tau_{ij}}\left(W_{ij} - B_i^{(j)}(ut_n-\tau_{ij})e^{-X_{ij}(ut_n-\tau_{ij})}\right)\right)^2$$
$$= T_1(n) + T_2(n).$$

We will analyze the latter term first. Since the summands are mean zero and independent it suffices to study the mean of the squares, i.e.

$$\mathbb{E}\left[e^{2ut_n(X_{1j}-\lambda_1)}e^{-2X_{1j}\tau_{1j}}\left(W_{1j} - B_1^{(j)}(ut_n-\tau_{1j})e^{-X_{1j}(ut_n-\tau_{ij})}\right)^2\right]$$
$$= \mathbb{E}\left[e^{2ut_n(X_{1j}-\lambda_1)}e^{-2X_{1j}\tau_{1j}}\mathbb{E}\left[\left(W_{1j} - B_1^{(j)}(ut_n-\tau_{1j})e^{-X_{1j}(ut_n-\tau_{ij})}\right)^2 |\tau_{1j}, X_{1j}\right]\right]$$
$$\leq \mathbb{E}\left[e^{2ut_n(X_{1j}-\lambda_1)}e^{-2X_{1j}\tau_{1j}}\kappa(X_{1j})e^{-X_{1j}(ut_n-\tau_{1j})}\right]$$
$$\leq Ke^{-\lambda_1 ut_n},$$

where the first inequality is due to (20). Since $E[N_j(ut_n)] = O(n^{-\alpha})$ it follows that $\lim_{n\to\infty} E[T_2(n)] = 0$.

Now consider $E[T_1(n); |Y_n^{(j)}(v)| > \varepsilon]$, and recalling that $E[N_j(ut_n)] = O(n^{-\alpha})$ observe that it suffices to show that each of the following goes to zero

$$E[W_{1j}^2; |Y_n^{(j)}(v)| > \varepsilon]\log n$$
$$\left(n^{-\lambda_1 u/r}\phi_1^{(j)}(ut_n)\right)^2 P(|Y_n^{(j)}(v)| > \varepsilon)\log n.$$

For the analysis of both of these terms it is useful to observe that $P(|Y_n^{(j)}(v)| > \varepsilon) \leq 1/(\varepsilon n)$. Next recall that W_{1j} inherits all of the moments of the offspring distribution, in particular $E[W_{1j}^4] < \infty$ in the case that the offspring distribution has a finite fourth moment (see e.g. the functional equation in (5) of I.6 of [2]). Therefore

$$E[W_{1j}^2; |Y_n^{(j)}(v)| > \varepsilon] \log n \leq (E[W_{1j}^4])^{1/2} \frac{\log n}{\varepsilon n^{1/2}},$$

which clearly goes to zero in the large n limit. Since

$$n^{-\lambda_1 u/r} \phi_1^{(j)}(ut_n) = n^{-\alpha} \int_0^{\lambda_1} g(x) e^{ut_n(x-\lambda_1)} \int_0^{ut_n} e^{-s(r+x)} ds dx = O\left(\frac{1}{n^\alpha \log n}\right)$$

it follows immediately that

$$\lim_{n\to\infty} \left(n^{-\lambda_1 u/r} \phi_1^{(j)}(ut_n)\right)^2 P(|Y_n^{(j)}(v)| > \varepsilon) \log^2 n = 0.$$

□

4.2.2 Proof of Lemma 2

Proof. First observe that for all $u \in [a,b]$ we have that $\sup_{n\geq 1} E[Y_n(u)^2] < \infty$, and hence for each u $\{Y_n(u)\}$ is a uniformly integrable sequence. Then consider the decomposition

$$\begin{aligned} Y_n(u) &= n^{(\alpha-1)/2}\left(e^{-\lambda_1 ut_n} Z_1(ut_n) - \frac{u}{n^\alpha}\int_0^{ut_n} Z_0(s) e^{-\lambda_1 s} ds\right) \\ &+ n^{-(1+\alpha)/2}\mu \int_0^{ut_n} e^{-\lambda_1 s}(Z_0(s) - ne^{-rs}) ds \\ &= M_n(u) + \mathcal{E}_n(u). \end{aligned} \tag{21}$$

Recalling that $\mathrm{Cov}(Z_0(s), Z_0(y)) = O(ne^{-rs})$, we have that

$$\mathbb{E}[\mathcal{E}_n(u)^2] = \frac{\mu}{n^{1+\alpha}} \int_0^{ut_n}\int_0^{ut_n} e^{-\lambda_1 s} \mathrm{Cov}(Z_0(s), Z_0(y)) ds dy = O(n^{-\alpha}),$$

which implies that $\sup_{n\geq 1} \mathbb{E}[M_n(u)^2] < \infty$. Furthermore the previous display implies that $\sup_{u\in[a,b]} \mathcal{E}_n(u) \to 0$ in probability, and hence it suffices to prove weak convergence for the sequence $\{M_n\}$.

Since for each n, M_n is a martingale it is possible to use the result of [11] to establish weak convergence. Specifically in order to establish this we need the FDD convergence and uniform integrability of $\{M_n\}$, as well as establish that the limit process Y satisfies property A from [11]. Since in Case I, Y is a constant process it is trivial to establish that the property holds.

□

4.2.3 Proof of Lemma 3

Proof. For $t > 0$ define $Z_{1,j}(t)$ to be all cells in population $Z_1(t)$ that are descended from mutations with fitness in $[x_{j-1}, x_j)$, which gives

$$Z_1(t) = \sum_{j=1}^{\lambda_1(\triangle_n)} Z_{1,j}(t).$$

Next consider the decomposition

$$\begin{aligned}\tilde{Y}_n(u) &= \left(n^{\alpha-1}t_n\right)^{1/2} e^{-\lambda_1 u t_n} \sum_{j=1}^{\lambda_1(\triangle_n)} \left(Z_{1,j}(ut_n) - \mu n^{1-\alpha} g_j \int_0^{ut_n} e^{-rs} e^{x_j(ut_n - s)} ds \right) \\ &+ \mu \left(n^{1-\alpha}t_n\right)^{1/2} \sum_{j=1}^{\lambda_1(\triangle_n)} \int_{x_{j-1}}^{x_j} \int_0^{ut_n} g(x) e^{-\lambda_1 u t_n - rs} \left(e^{x(ut_n - s)} - e^{x_j(ut_n - s)} \right) ds dx \\ &\doteq I_{n,1}(u) + I_{n,2}(u).\end{aligned}$$

Note that by continuity we have that $\sup_{u\in[a,b]} |I_{n,2}(u)| = O(\triangle_n n^{(1-\alpha)/2} t_n^{1/2})$, which goes to zero by our choice of $\triangle_n$.

For what follows it is convenient to re-organize $I_{n,1}(u)$ as follows

$$I_{n,1}(u) = \left(n^{\alpha-1}t_n\right)^{1/2} \sum_{j=1}^{\lambda_1(\triangle_n)} e^{ut_n(x_j - \lambda_1)} \left(Z_{1,j}(ut_n) e^{-ut_n x_j} - \frac{\mu g_j}{n^{\alpha-1}} \int_0^{ut_n} e^{-s(r+x_j)} ds \right). \tag{22}$$

Note that $Z_{1,j}(\cdot)$ is a binary branching process whose cells have birth rate in $[d_0 + x_{j-1}, d_0 + x_j]$, death rate d_0 and immigration rate $\mu n^{-\alpha} Z_0(t) \int_{x_{j-1}}^{x_j} g(x)dx$. We will now create a family of coupled processes $\{\hat{Z}_{1,j}(\cdot)\}_{j=1}^{\lambda_1(\triangle_n)}$ such that for each j and $t \geq 0$, the inequality $Z_{1,j}(t) \leq \hat{Z}_{1,j}(t)$ holds a.s.

For a new resistant cell A created in the $Z_{1,j}$ population with birth rate $d_0 + x$, create a matched cell A^* with birth rate $d_0 + x_j$. Each birth and death by A and its descendants is matched by a birth or death by A^* or its descendants. In addition to matching A and its descendants, the A^* cells will have additional births at rate $x_j - x$. One offspring from this event will continue tracking the behavior of the A cell. The other offspring initiates a new branching process with death rate d_0 and birth rate $d_0 + x_j$. The total population of A^* cells and their descendants comprise the population of the process $\hat{Z}_{1,j}$. Note that $\hat{Z}_{1,j}$ is a branching process with birth rate $d_0 + x_j$, death rate d_0, and immigration rate at time s, $Z_0(s)\mu n^{-\alpha} \int_{x_{j-1}}^{x_j} g(x)dx$.

For $t \geq 0$ let $N_j(t)$ be the number of mutations with fitness in $[x_{j-1}, x_j)$ that have occurred by time t. Enumerate the fitness of the mutants by $\{x_j^{(i)}\}_{i\geq 1}$. Note then that

$$\hat{Z}_{1,j}(t) - Z_{1,j}(t) = \sum_{i=1}^{N_j(t)} B_j^{(i)}(t),$$

where $B_j^{(i)}$ is a branching process with net growth rate x_j and immigration at rate $(x_j - x_j^{(i)})Z_{1,j}^{(i)}(s)$, where $Z_{1,j}^{(i)}$ is the contribution of the ith mutation to $Z_{1,j}$. Note that we can create another family of coupled processes W_j such that for each $t \geq 0$

$$\sum_{i=1}^{N_j(t)} B_j^{(i)}(t) \leq W_j$$

where W_j is a binary branching process with birth rate $d_1 + x_j$, death rate d_1 and immigration at rate $\triangle_n \hat{Z}_{1,j}$. Thus for $t \geq 0$ and each $j \in \{0, \ldots, \lambda_1(\triangle)\}$

$$0 \leq \hat{Z}_{1,j}(t) - Z_{1,j}(t) \leq W_j(t).$$

Returning to $I_{n,1}(u)$ we have that

$$\begin{aligned}
I_{n,1}(u) &= \left(n^{\alpha-1}t_n\right)^{1/2} \sum_{j=1}^{\lambda_1(\triangle_n)} e^{ut_n(x_j-\lambda_1)} \left(\hat{Z}_{1,j}(ut_n)e^{-ut_nx_j} - \frac{\mu g_j}{n^{\alpha-1}} \int_0^{ut_n} e^{-s(r+x_j)}ds\right) \\
&\quad + \left(n^{\alpha-1}t_n\right)^{1/2} \sum_{j=1}^{\lambda_1(\triangle_n)} e^{-\lambda_1 ut_n} \left(Z_{1,j}(ut_n) - \hat{Z}_{1,j}(ut_n)\right) \\
&= \hat{I}_{n,1}(u) + I_{n,3}(u).
\end{aligned}$$

We would like to convert the summands in $\hat{I}_{n,1}(u)$ into martingales. In order to do this we need to replace ne^{-rs} with $Z_0(s)$, which gives

$$\begin{aligned}
\hat{I}_{n,1}(u) &= \left(n^{\alpha-1}t_n\right)^{1/2} \sum_{j=1}^{\lambda_1(\triangle_n)} e^{ut_n(x_j-\lambda_1)} \left(\hat{Z}_{1,j}(ut_n)e^{-ut_nx_j} - \frac{\mu g_j}{n^{\alpha}} \int_0^{ut_n} e^{-sx_j} Z_0(s)ds\right) \\
&\quad + \mu \left(\frac{t_n}{n^{\alpha+1}}\right)^{1/2} \sum_{j=1}^{\lambda_1(\triangle_n)} e^{ut_n(x_j-\lambda_1)} g_j \int_0^{ut_n} e^{-sx_j} \left(Z_0(s) - ne^{-rs}\right) ds \\
&= J_n(u) + I_{n,4}(u).
\end{aligned}$$

We will now show that for our choice of $\triangle_n$ for any $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\sup_{u\in[a,b]} |I_{n,k}(u)| > \varepsilon\right) = 0$$

for $k = 2, 3, 4$. The result for $I_{n,2}$ has already been established, so now consider $I_{n,3}$:

$$\begin{aligned}
\sup_{u\in[a,b]} |I_{n,3}(u)| &\leq \left(n^{\alpha-1}t_n\right)^{1/2} \sup_{u\in[a,b]} e^{-\lambda_1 ut_n} \sum_{j=1}^{\lambda_1(\triangle_n)} \left|\hat{Z}_{1,j}(ut_n) - Z_{1,j}(ut_n)\right| \\
&\leq \left(n^{\alpha-1}t_n\right)^{1/2} e^{-\lambda_1 at_n} \sup_{u\in[a,b]} \sum_{j=1}^{\lambda_1(\triangle_n)} W_j(ut_n).
\end{aligned}$$

Note that $\sum_{j=1}^{\lambda_1(\triangle_n)} W_j(ut_n)$ is a non-negative submartingale with respect to the filtration generated by $\{W_j, \hat{Z}_{1,j}\}_{j=1}^{\lambda_1(\triangle_n)}$. Thus from Doob's inequality we have that

$$P\left(\sup_{u\in[a,b]} \sum_{j=1}^{\lambda_1(\triangle_n)} W_j(ut_n) > \varepsilon\right) \le \frac{1}{\varepsilon}\sum_{j=1}^{\lambda_1(\triangle_n)} E[W_j(bt_n)]$$
$$= \frac{\triangle_n}{\varepsilon}\sum_{j=1}^{\lambda_1(\triangle_n)} \int_0^{bt_n} e^{x_j(bt_n-s)} E[\hat{Z}_{1,j}(s)]ds.$$

For $s \ge 0$ we have that

$$E[\hat{Z}_{1,j}(s)] = \frac{\mu g_j}{n^\alpha}\int_0^s E[Z_0(y)]e^{x_j(s-y)}dy \le \frac{e^{x_j s}\mu g_j n^{1-\alpha}}{r+x_j},$$

and therefore

$$\sum_{j=1}^{\lambda_1(\triangle_n)} E[W_j(ut_n)] = O\left(\triangle_n n^{1-\alpha} e^{\lambda_1 bt_n} t_n\right).$$

Thus the desired result will follow by our choice of $\triangle_n$.

We next consider

$$\sup_{u\in[a,b]} |I_{n,4}(u)| \le \left(\frac{t_n}{n^{1+\alpha}}\right)^{1/2} \sum_{j=1}^{\lambda_1(\triangle_n)} e^{at_n(x_j-\lambda_1)} g_j \int_0^{bt_n} |Z_0(s) - ne^{-rs}|ds$$
$$\le \left(\frac{t_n}{n^{1+\alpha}}\right)^{1/2} \int_0^{bt_n} |Z_0(s) - ne^{-rs}|ds,$$

where we get the inequality via $e^{at_n(x_j-\lambda_1)} \le 1$. Since $\mathrm{Var}(Z_0(s)) = O(e^{-rs}n)$ it follows from the Chebyshev and Cauchy-Schwarz inequality that

$$P\left(\sup_{u\in[a,b]} |I_{n,4}(u)| > \varepsilon\right) \le \frac{1}{\varepsilon} O(t_n/n^\alpha)^{1/2}.$$

□

4.2.4 Proof of Lemma 4

Proof. For each $j \in \{1,\ldots,\lambda_1(\triangle_n)\}$ define for $u \in [a,b]$

$$M_j(ut_n) = \left(n^{\alpha-1}t_n\right)^{1/2}\left(\hat{Z}_{1,j}(ut_n)e^{-ut_n x_j} - \frac{\mu g_j}{n^\alpha}\int_0^{ut_n} Z_0(s)e^{-sx_j}ds\right),$$

note that $M_j(ut_n)$ is a martingale (in u) with respect to the filtration

$$\hat{\mathcal{F}}_t = \sigma\left(\hat{Z}_{1,1}(t),\ldots,\hat{Z}_{1,\lambda_1(\triangle_n)}(t), Z_0(t), t \le ut_n\right),$$

and that

$$J_n(u) = \sum_{j=1}^{\lambda_1(\triangle_n)} e^{ut_n(x_j-\lambda_1)} M_j(ut_n).$$

In order to establish the tightness of J_n we will establish that there exists a $C > 0$ and $\beta > 1$ such that for $u \in [a,b]$ and $0 \le h \le u-a$

$$E\left[(J_n(u+h)-J_n(u))^2 (J_n(u)-J_n(u-h))^2\right] \le Ch^\beta, \tag{23}$$

see e.g. Theorem 10.4 or 13.5 of [3]. First calculate

$$(J_n(u+h)-J_n(u))^2 = \sum_{j=1}^{\lambda_1(\triangle_n)} \left(e^{(u+h)(x_j-\lambda_1)t_n} M_j((u+h)t_n) - e^{ut_n(x_j-\lambda_1)} M_j(ut_n)\right)^2$$

$$+ \sum_{j=1}^{\lambda_1(\triangle_n)} \sum_{i\neq j} e^{ut_n(x_i+x_j-2\lambda_1)} \left(e^{ht_n(x_i-\lambda_1)} M_i((u+h)t_n) - M_i(ut_n)\right) \left(e^{ht_n(x_j-\lambda_1)} M_j((u+h)t_n) - M_j(ut_n)\right).$$

Next consider the sigma algebra $\mathcal{F}^0_\infty = \sigma(Z_0(s), s \ge 0)$ and the enlarged filtration $\tilde{\mathcal{F}}_{ut_n} = \sigma\left(\hat{\mathcal{F}}_{ut_n} \cup \mathcal{F}^0_\infty\right)$. Since M_j is still a martingale with respect to $\tilde{\mathcal{F}}_{ut_n}$, and the processes M_j and M_i are independent conditional on $\mathcal{F}^0_\infty$ we can use the tower property to see that

$$E\left[M_i((u+h)t_n) M_j((u+h)t_n) | \hat{\mathcal{F}}_{ut_n}\right] = M_i(ut_n) M_j(ut_n).$$

Then based on the previous display we have that

$$\begin{aligned} &E\left[(J_n(u+h)-J_n(u))^2 | \hat{\mathcal{F}}_{ut_n}\right] \\ &= \sum_{j=1}^{\lambda_1(\triangle_n)} E\left[\left(e^{(u+h)(x_j-\lambda_1)t_n} M_j((u+h)t_n) - e^{ut_n(x_j-\lambda_1)} M_j(ut_n)\right)^2 | \hat{\mathcal{F}}_{ut_n}\right] \\ &+ \sum_{j=1}^{\lambda_1(\triangle_n)} \sum_{i\neq j} M_i(ut_n) M_j(ut_n) e^{ut_n(x_i+x_j-2\lambda_1)} \left(e^{ht_n(x_i-\lambda_1)}-1\right)\left(e^{ht_n(x_j-\lambda_1)}-1\right). \end{aligned} \tag{24}$$

We can use the martingale property to see that

$$\begin{aligned} &E\left[\left(e^{(u+h)(x_j-\lambda_1)t_n} M_j((u+h)t_n) - e^{ut_n(x_j-\lambda_1)} M_j(ut_n)\right)^2 | \hat{\mathcal{F}}_{ut_n}\right] \\ &= M_j(ut_n)^2 e^{2ut_n(x_j-\lambda_1)} \left(e^{ht_n(x_j-\lambda_1)}-1\right)^2 \\ &\quad + e^{2(u+h)(x_j-\lambda_1)t_n} \left(E\left[M_j((u+h)t_n)^2 | \hat{\mathcal{F}}_{ut_n}\right] - M_j(ut_n)^2\right). \end{aligned} \tag{25}$$

We can then calculate that

$$
\begin{aligned}
&\frac{n^{1-\alpha}}{t_n}\left(E\left[M_j\left((u+h)t_n\right)^2|\hat{\mathcal{F}}_{ut_n}\right]-M_j(ut_n)^2\right)=e^{-2t_nx_j(u+h)}E\left[\hat{Z}_{1,j}\left((u+h)t_n\right)^2|\hat{\mathcal{F}}_{ut_n}\right]\\
&\quad -e^{-2ux_jt_n}\hat{Z}_{1,j}(ut_n)^2+2\left(\frac{\mu g_j}{n^\alpha}\right)^2\int_{ut_n}^{(u+h)t_n}\int_0^{(u+h)t_n}e^{-x_j(s+y)}E[Z_0(s)Z_0(y)|\hat{\mathcal{F}}_{ut_n}]dsdy\\
&\quad -\left(\frac{\mu g_j}{n^\alpha}\right)^2\int_{ut_n}^{(u+h)t_n}\int_{ut_n}^{(u+h)t_n}e^{-x_j(s+y)}E[Z_0(s)Z_0(y)|\hat{\mathcal{F}}_{ut_n}]dsdy\\
&\quad +\frac{2\mu g_je^{-ux_jt_n}}{n^\alpha}\int_0^{ut_n}e^{-sx_j}\left(Z_0(s)\hat{Z}_{1,j}(ut_n)-e^{-x_jht_n}E[Z_0(s)\hat{Z}_{1,j}\left((u+h)t_n\right)|\hat{\mathcal{F}}_{ut_n}]\right)ds\\
&\quad -\frac{2\mu g_je^{-(u+h)x_jt_n}}{n^\alpha}\int_{ut_n}^{(u+h)t_n}e^{-sx_j}E[Z_0(s)\hat{Z}_{1,j}\left((u+h)t_n\right)|\hat{\mathcal{F}}_{ut_n}]ds,
\end{aligned}
$$

we can evaluate the penultimate term in the previous display and simplify to get that

$$
\begin{aligned}
&\frac{n^{1-\alpha}}{t_n}\left(E\left[M_j\left((u+h)t_n\right)^2|\hat{\mathcal{F}}_{ut_n}\right]-M_j(ut_n)^2\right)\\
&=e^{-2t_nx_j(u+h)}E\left[\hat{Z}_{1,j}\left((u+h)t_n\right)^2|\hat{\mathcal{F}}_{ut_n}\right]\\
&\quad -e^{-2ux_jt_n}\hat{Z}_{1,j}(ut_n)^2-\frac{2\mu g_je^{-ux_jt_n}}{n^\alpha}\hat{Z}_{1,j}(ut_n)\int_{ut_n}^{(u+h)t_n}E[Z_0(s)|\hat{\mathcal{F}}_{ut_n}]ds\\
&\quad -\left(\frac{\mu g_j}{n^\alpha}\right)^2\int_{ut_n}^{(u+h)t_n}\int_{ut_n}^{(u+h)t_n}e^{-x_j(s+y)}E[Z_0(s)Z_0(y)|\hat{\mathcal{F}}_{ut_n}]dsdy. \qquad (26)
\end{aligned}
$$

For each j, let $\tilde{Z}_{1,j}$ denote a binary branching process with birth rate d_0+x_j and death rate x_j and initial condition 1. We can now calculate that

$$
\begin{aligned}
&e^{-2t_nx_j(u+h)}E\left[\hat{Z}_{1,j}\left((u+h)t_n\right)^2|\hat{\mathcal{F}}_{ut_n}\right]=e^{-2ut_nx_j}\left(\hat{Z}_{1,j}(ut_n)^2+e^{-2ht_nx_j}\hat{Z}_{1,j}(ut_n)\mathrm{Var}\left(\tilde{Z}_{1,j}(ht_n)\right)\right)\\
&\quad +\frac{2\mu g_j\hat{Z}_{1,j}(ut_n)}{n^\alpha e^{ut_nx_j}}\int_{ut_n}^{(u+h)t_n}e^{-x_js}E[Z_0(s)|\hat{\mathcal{F}}_{ut_n}]ds\\
&\quad +\frac{\mu}{n^\alpha e^{-2t_nx_j(u+h)}}\int_{ut_n}^{(u+h)t_n}E[Z_0(s)|\hat{\mathcal{F}}_{ut_n}]E\left[\tilde{Z}_{1,j}\left((u+h)t_n-s\right)^2\right]ds\\
&\quad +\left(\frac{\mu g_j}{n^\alpha}\right)^2\int_{ut_n}^{(u+h)t_n}\int_{ut_n}^{(u+h)t_n}E[Z_0(s)Z_0(y)|\hat{\mathcal{F}}_{ut_n}]e^{-x_j(s+y)}dsdy.
\end{aligned}
$$

Using the previous display we can simplify (26) to the following form

$$
\begin{aligned}
&E\left[M_j\left((u+h)t_n\right)^2|\hat{\mathcal{F}}_{ut_n}\right]-M_j(ut_n)^2=e^{-2t_nx_j(u+h)}n^{\alpha-1}t_n\hat{Z}_{1,j}(ut_n)\mathrm{Var}\left(\tilde{Z}_{1,j}(ht_n)\right)\\
&\quad +\frac{\mu g_jt_n}{ne^{2t_nx_j(u+h)}}\int_{ut_n}^{(u+h)t_n}E[Z_0(s)|\hat{\mathcal{F}}_{ut_n}]E\left[\tilde{Z}_{1,j}((u+h)t_n-s)^2\right]ds.
\end{aligned}
$$

For ease of notation we define

$$H_j(u,h;n) = E\left[M_j\left((u+h)t_n\right)^2 | \hat{\mathcal{F}}_{ut_n}\right] - M_j(ut_n)^2 \tag{27}$$

We can then plug (25) into (24) to get that

$$\begin{aligned} & E\left[\left(J_n(u+h) - J_n(u)\right)^2 | \hat{\mathcal{F}}_{ut_n}\right] \qquad (28) \\ & \leq \sum_{j=1}^{\lambda_1(\triangle_n)} M_j(ut_n)^2 e^{2ut_n(x_j-\lambda_1)} \left(e^{ht_n(x_j-\lambda_1)} - 1\right)^2 + \sum_{j=1}^{\lambda_1(\triangle_n)} e^{2(u+h)(x_j-\lambda_1)t_n} H_j(u,h;n). \\ & \quad + \sum_{j=1}^{\lambda_1(\triangle_n)} \sum_{i\neq j} M_i(ut_n) M_j(ut_n) e^{ut_n(x_i+x_j-2\lambda_1)} \left(e^{ht_n(x_i-\lambda_1)} - 1\right)\left(e^{ht_n(x_j-\lambda_1)} - 1\right). \end{aligned}$$

Noting that if $i \neq j$ and $k = 1$ or $\ell = 1$ then $E[M_i(ut_n)^k M_j(vt_n)^\ell | \mathcal{F}_\infty^0] = 0$, we can write

$$\begin{aligned} & E\left[\left(J(u) - J(u-h)\right)^2 E\left[\left(J(u+h) - J(u)\right)^2 | \hat{\mathcal{F}}_{ut_n}\right]\right] \\ & = \sum_{i=1}^{\lambda_1(\triangle_n)} \sum_{j\neq i} e^{2ut_n(x_j+x_i-2\lambda_1)} E\left[M_i(ut_n) M_j(ut_n)\left(M_j(ut_n) - e^{-ht_n(x_j-\lambda_1)} M_j((u-h)t_n)\right)\right. \\ & \quad \left. \times \left(M_i(ut_n) - e^{-ht_n(x_i-\lambda_1)} M_i((u-h)t_n)\right)\right]\left(e^{ht_n(x_j-\lambda_1)} - 1\right)\left(e^{ht_n(x_i-\lambda_1)} - 1\right) \\ & \quad + \sum_{j=1}^{\lambda_1(\triangle_n)} e^{4ut_n(x_j-\lambda_1)} \left(e^{ht_n(x_j-\lambda_1)} - 1\right)^2 E\left[M_j(ut_n)^2 \left(M_j(ut_n) - e^{-ht_n(x_j-\lambda_1)} M_j((u-h)t_n)\right)^2\right] \\ & \quad + \sum_{i=1}^{\lambda_1(\triangle_n)} \sum_{j\neq i} e^{2ut_n(x_j+x_i-2\lambda_1)} \left(e^{ht_n(x_j-\lambda_1)} - 1\right)^2 E\left[M_j(ut_n)^2 \left(M_i(ut_n) - e^{-ht_n(x_i-\lambda_1)} M_i((u-h)t_n)\right)^2\right] \\ & \quad + \sum_{j=1}^{\lambda_1(\triangle_n)} e^{2t_n(2u+h)(x_j-\lambda_1)} E\left[H_j(u,h;n)\left(M_j(ut_n) - e^{-ht_n(x_j-\lambda_1)} M_j((u-h)t_n)\right)^2\right] \\ & \quad + \sum_{i=1}^{\lambda_1(\triangle_n)} \sum_{j\neq i} e^{2(u+h)(x_i-\lambda_1)t_n} e^{2ut_n(x_j-\lambda_1)} E\left[H_i(u,h;n)\left(M_j(ut_n) - e^{-ht_n(x_j-\lambda_1)} M_j((u-h)t_n)\right)^2\right] \\ & = L_1(u,h;n) + L_2(u,h;n) + L_3(u,h;n) + L_4(u,h;n) + L_5(u,h;n). \end{aligned}$$

Thus if we establish that there is a $C > 0$ and $\beta > 1$ such that for all $1 \leq m \leq 5$, $u \in [a,b]$ and $0 \leq h \leq u - a$ we have $L_m(u,h;n) \leq Ch^\beta$ then we will establish the tightness condition (23). Each of the terms L_m can be written as either a single sum of the form

$$S_m(u,h;n) = \sum_{j=1}^{\lambda_1(\triangle_n)} e^{k_m ut_n(x_j-\lambda_1)} E[\Psi_{j,m}(u,h;n)],$$

for appropriate random variable $\Psi_{j,m}$ and non-negative integer k_m or alternatively as the double sum

$$D_m(u,h;n) = \sum_{j=1}^{\lambda_1(\triangle_n)} \sum_{i \neq j} e^{k_m u t_n (x_j - \lambda_1)} e^{\ell_m u t_n (x_i - \lambda_1)} E[\Psi_{j,m}(u,h;n) \Phi_{i,m}(u,h;n)].$$

Claim A: Suppose that for each $j \in \{1, \ldots, \lambda_1(\triangle_n)\}$ there exists a function $\rho_j(h;n)$ such that

$$e^{k_m u t_n (x_j - \lambda_1)} \rho_j(h;n) \leq K_1 e^{v t_n (x_j - \lambda_1)} \quad (A)$$

for positive constants K_1, v independent of j, h and n. If the following bound is satisfied

$$E[\Psi_{j,m}(u,h;n)] \leq C_0 h^\beta g_j t_n \left(t_n(\lambda_1 - x_j)\right)^{\kappa_m} \rho_j(h;n) \tag{29}$$

for non-negative integer κ_m, then $S_m(u,h;n) \leq C h^\beta$.
Proof: We prove the claim for the single sum. Thus consider

$$\begin{aligned}
S_m(u,h;n) &\leq C_0 h^\beta t_n \sum_{j=1}^{\lambda_1(\triangle_n)} e^{k_m u t_n (x_j - \lambda_1)} g_j \left(t_n(\lambda_1 - x_j)\right)^{\kappa_m} \rho_j(h;n) \\
&\leq C_0 h^\beta t_n \sum_{j=1}^{\lambda_1(\triangle_n)} e^{v t_n (x_j - \lambda_1)} \int_{x_{j-1}}^{x_j} \left(t_n(\lambda_1 - x)\right)^{\kappa_m} g(x) dx \\
&\leq C_0 h^\beta t_n e^{u t_n \triangle_n} \int_0^{\lambda_1} e^{v t_n (x - \lambda_1)} \left(t_n(\lambda_1 - x)\right)^{\kappa_m} g(x) dx \\
&\leq C_1 h^\beta \int_0^{\lambda_1 t_n} e^{-vy} y^{\kappa_m} g(\lambda_1 - y/t_n) dy \leq C h^\beta,
\end{aligned}$$

where the penultimate inequality follows by the requirement that $\sup_n t_n \triangle_n < \infty$, and an application of the change of variable $y = t_n(\lambda_1 - x)$, and the final inequality follows from the assumption that the density g is bounded. ■

In order to establish a corresponding result for the double sum terms note that conditioned on $\mathcal{F}_\infty^0 = \sigma(Z_0(s), s \geq 0)$ the random variables Ψ_j and Φ_i will be independent. Thus for the double sum it suffices to show that there there exist two functions $\rho_j^1(h;n)$, $\rho_j^2(h;n)$ for each $j \in \{1, \ldots, \lambda_1(\triangle_n)\}$ satisfying (A) from Claim A, and such that

$$E[\Psi_{j,m}(u,h;n) | \mathcal{F}_\infty^0] \leq C_0 h^{\gamma_1} g_j t_n \left(t_n(\lambda_1 - x_j)\right)^{\kappa_m} \rho_j^1(h;n) \left(\frac{c_1 Z_0(u t_n)}{n} + c_2 \int_0^{u t_n} \frac{Z_0(s)}{n} ds \right) \tag{30}$$

$$E[\Phi_{j,m}(u,h;n) | \mathcal{F}_\infty^0] \leq C_0 h^{\gamma_2} g_j t_n \left(t_n(\lambda_1 - x_j)\right)^{\ell_m} \rho_j^2(h;n) \left(\frac{c_3 Z_0(u t_n)}{n} + c_4 \int_0^{u t_n} \frac{Z_0(s)}{n} ds \right)$$

for non-negative integers κ_m, ℓ_m, non-negative constants c_1, c_2, c_3, c_4, and $\gamma_1 + \gamma_2 = \beta$. If the above holds, we can establish that $D_m(u, h; n) \leq Ch^\beta$ by noting that $E[(Z_0(ut_n)/n)^2] = O(1)$ and that

$$(1/n)^2 \int_0^{ut_n} \int_0^{ut_n} E[Z_0(s)Z_0(y)]dsdy = O(1).$$

We now will verify (29) or (30) for each $1 \leq m \leq 5$.

First consider L_1, which is a double sum with $\Psi_{j,1} = \Phi_{j,1}$ and

$$\begin{aligned}\Psi_{j,1}(u, h; n) &= \left(e^{ht_n(\lambda_1 - x_j)} - 1\right) \left(M_j(ut_n)^2 - e^{ht_n(\lambda_1 - x_j)} M_j(ut_n) M_j((u-h)t_n)\right) \\ &\leq ht_n(\lambda_1 - x_j) \left(M_j(ut_n)^2 - e^{ht_n(\lambda_1 - x_j)} M_j(ut_n) M_j((u-h)t_n)\right).\end{aligned}$$

If we define

$$\begin{aligned}\tilde{H}_j(u, h; n) &= E\left[M_j\left((u+h)t_n\right)^2 | \tilde{\mathcal{F}}_{ut_n}\right] - M_j(ut_n)^2 \qquad (31) \\ &= e^{-2t_n x_j(u+h)} n^{\alpha-1} t_n \hat{Z}_{1,j}(ut_n) \mathrm{Var}\left(\tilde{Z}_{1,j}(ht_n)\right) \\ &\quad + \frac{\mu g_j t_n}{n e^{2t_n x_j(u+h)}} \int_{ut_n}^{(u+h)t_n} Z_0(s) E\left[\tilde{Z}_{1,j}((u+h)t_n - s)^2\right] ds.\end{aligned}$$

it then follows from the martingale property that

$$\begin{aligned}E[\Psi_{j,1}(u, h; n)|\mathcal{F}_\infty^0] &\leq ht_n(\lambda_1 - x_j) E\left[E\left[M_j(ut_n)^2 | \tilde{\mathcal{F}}_{(u-h)t_n}\right] - e^{ht_n(\lambda_1 - x_j)} M_j((u-h)t_n)^2 | \mathcal{F}_\infty^0\right] \\ &= ht_n(\lambda_1 - x_j) E\left[\tilde{H}_j(u-h, h; n) + M_j((u-h)t_n)^2 (1 - e^{ht_n(\lambda_1 - x_j)}) | \mathcal{F}_\infty^0\right] \\ &\leq ht_n(\lambda_1 - x_j) E\left[\tilde{H}_j(u-h, h; n) | \mathcal{F}_\infty^0\right]\end{aligned}$$

Using the formulas for first and second moments of Markovian branching processes one can establish that there are constants a_1, a_2 such that

$$\begin{aligned}E[\tilde{H}_j(u-h, h; n)|\mathcal{F}_\infty^0] &\leq \frac{a_1 \mu g_j t_n e^{-t_n x_j(u-h)}}{n} \int_0^{ht_n} e^{-x_j s} ds \int_0^{(u-h)t_n} Z_0(s) e^{-x_j s} ds \qquad (32) \\ &\quad + \frac{a_2 \mu g_j t_n}{n} \int_{(u-h)t_n}^{ut_n} Z_0(s) ds \\ &\leq \frac{a_3 g_j t_n m_j(h; n)}{n} \int_0^{ut_n} Z_0(s) ds\end{aligned}$$

where $a_3 = 2\mu * \max(a_1, a_2)$ and $m_j(h; n) = 2\max(1, (1 - e^{-x_j h t_n})/x_j) \leq 2\max(1, \min(ht_n, 1/x_j))$. Thus to verify the conditions of Claim A we must verify that $m_j(h; n)$ satisfies condition (A). Note that if $j \geq \lambda_1(\triangle_n)/2$ then $m_j(h; n) \leq 2\max(1, 2/\lambda_1)$, while for $j \leq \lambda_1(\triangle_n)/2$ and n sufficiently large $m_j(h; n) \leq ht_n$, and $e^{ut_n(\lambda_1 - x_j)/2} \geq e^{\lambda_1 ut_n/4}$. Thus we have $m_j(h; n) \leq (4/\lambda_1) e^{ut_n(\lambda_1 - x_j)/4}$, satisfying condition (A) of Claim A.

We now consider the L_2 term, which is a single sum with

$$\begin{aligned}\Psi_{j,2}(u,h;n) &= \left(e^{ht_n(x_j-\lambda_1)}-1\right)^2 E\left[M_j(ut_n)^2\left(M_j(ut_n)-e^{-ht_n(x_j-\lambda_1)}M_j((u-h)t_n)\right)^2\right] \\ &\leq (ht_n(\lambda_1-x_j))^2 E\left[M_j(ut_n)^2\left(M_j(ut_n)-e^{-ht_n(x_j-\lambda_1)}M_j((u-h)t_n)\right)^2\right].\end{aligned}$$

Applying Hölder's inequality and observing that M_j^4 is a submartingale we observe that

$$E\left[M_j(ut_n)^2\left(M_j(ut_n)-e^{-ht_n(x_j-\lambda_1)}M_j((u-h)t_n)\right)^2\right] \leq 3e^{2ht_n(\lambda_1-x_j)}E[M_j(ut_n)^4].$$

Thus in the language of Claim A we have that $\rho_j(h;n)=e^{2ht_n(\lambda_1-x_j)}$. Since $k_2=4$, and by definition $h\leq u$ we see that $e^{2ht_n(\lambda_1-x_j)}\leq e^{2ut_n(\lambda_1-x_j)}$, and hence $e^{2ht_n(\lambda_1-x_j)}$ satisfies condition (A) of Claim A. In order to verify that Claim A is applicable we thus need to establish that $E[M_j(ut_n)^4]\leq cg_jt_n$. Then observe that

$$E[M_j(ut_n)^4] \leq c(n^{\alpha-1}t_n)^2\left(e^{-4ut_nx_j}E[\hat{Z}_{1,j}(ut_n)^4]+\left(\frac{g_j}{n^\alpha}\right)^4 E\left(\int_0^{ut_n} Z_0(s)ds\right)^4\right). \tag{33}$$

Apply Jensen's inequality to see that

$$E\left(\int_0^{ut_n} Z_0(s)ds\right)^4 \leq (ut_n)^3\int_0^{ut_n} E[Z_0(s)^4]ds = O(n^4t_n^3), \tag{34}$$

and therefore the latter term in (33) is $O(g_j^4n^{2(1-\alpha)}t_n^5)$ and recalling that $g(x)\leq G$ for all x we see that if $\triangle_n^3=O(n^{2(\alpha-1)}/t_n^4)$ the latter term in (33) is $O(g_jt_n)$. Observe that

$$\hat{Z}_{1,j}(t)=\sum_{i=1}^{N_j(t)}\hat{B}_{i,j}(t-\tau_i),$$

where N_j is a Poisson process with intensity at time s given by $\lambda(s)=\mu g_jn^{-\alpha}Z_0(s)$, τ_i is the time of creation of the ith mutant and $\hat{B}_{i,j}$ is a binary branching process with birth rate d_0+x_j and death rate d_0 (note we denote a generic copy of the branching process by $\hat{B}_j$) Therefore if we define $\Lambda(t)=\int_0^t\lambda(s)ds$ and observe that $E[\hat{B}_j(t)^k]$ is increasing in t for positive integer k

$$\begin{aligned}E[\hat{Z}_{1,j}(ut_n)^4] \leq{}& E[\Lambda(ut_n)]E[\hat{B}_j(ut_n)^4]+3E[\Lambda(ut_n)^2](E[\hat{B}_j(ut_n)^2])^2 \\ &+E[\Lambda(ut_n)^4](E[\hat{B}_j(ut_n)])^4+4E[\Lambda(ut_n)^2]E[\hat{B}_j(ut_n)^3]E[\hat{B}_j(ut_n)] \\ &+6E[\Lambda(ut_n)^2]E[\hat{B}_j(ut_n)^2](E[\hat{B}_j(ut_n)])^2.\end{aligned}$$

From Lemma 5, we know that $E[\hat{B}_j(ut_n)^k]=O(e^{x_jut_nk})$ for $k\leq 4$. Using the same argument as we used above in (34) we see that$E[\Lambda(ut_n)^k]=O(g_jn^{k(1-\alpha)}t_n^{k-1}\triangle_n^{k-1})$

and thus there exists a $K > 0$ such that

$$
\begin{aligned}
&e^{-4ut_n x_j} n^{2(\alpha-1)} t_n^2 E[\hat{Z}_{1,j}(ut_n)^4] \\
&\leq K n^{2(\alpha-1)} t_n^2 \left(E[\Lambda(ut_n)] + 3E[\Lambda(ut_n)^2] + E[\Lambda(ut_n)^4] + 4E[\Lambda(ut_n)^2] + 6E[\Lambda(ut_n)^2] \right) \\
&\leq K g_j t_n^2 \left(n^{\alpha-1} + t_n \triangle_n + n^{2(1-\alpha)} t_n^3 \triangle_n^3 \right) = O(g_j t_n),
\end{aligned}
$$

where the final equality follows from our choice of $\triangle_n$.

We now consider the L_3 term, which is a double sum with

$$
\begin{aligned}
\Psi_{j,3}(u,h;n) &= \left(e^{ht_n(x_j-\lambda_1)} - 1 \right)^2 M_j(ut_n)^2 \\
\Phi_{i,3}(u,h;n) &= \left(M_i(ut_n) - e^{ht_n(\lambda_1 - x_j)} M_i\left((u-h)t_n\right) \right)^2.
\end{aligned}
$$

We first use the martingale property of M_i to calculate that

$$
\begin{aligned}
E[\Phi_{i,3}(u,h;n)|\mathcal{F}_\infty^0] &= E\left[E[M_i(ut_n)^2|\tilde{F}_{(u-h)t_n}] - M_i((u-h)t_n)^2|\mathcal{F}_\infty^0 \right] \\
&\quad + E[M_i((u-h)t_n)^2|\mathcal{F}_\infty^0] \left(e^{ht_n(\lambda_1-x_i)} - 1 \right)^2 \\
&= E[\tilde{H}_i(u-h,h;n)|\mathcal{F}_\infty^0] + E[M_i((u-h)t_n)^2|\mathcal{F}_\infty^0] \left(e^{ht_n(\lambda_1-x_i)} - 1 \right)^2.
\end{aligned} \tag{35}
$$

It is easy to see that $e^{-2ht_n x_j}\mathrm{Var}(\tilde{Z}_{1,j}(ht_n)) = O(ht_n)$ and that there exists a constant $C > 0$ such that for all j we have

$$
e^{-2t_n x_j(u+h)} E\left[\tilde{Z}_{1,j}\left((u+h)t_n - s\right) \right] \leq Ce^{-x_j s},
$$

and thus there exists positive constants C_1, C_2 such that

$$
\begin{aligned}
E\left[\tilde{H}_i(u-h,h;n)|\mathcal{F}_\infty^0 \right] &\leq \frac{C_1 t_n^2 g_i h}{n e^{t_n x_i(u-h)}} \int_0^{(u-h)t_n} Z_0(s) e^{-x_i s} ds \\
&\quad + \frac{C_2 g_i t_n}{n} \int_{(u-h)t_n}^{ut_n} Z_0(s) e^{-x_i s} ds.
\end{aligned}
$$

Furthermore we can then calculate that

$$
E[M_i((u-h)t_n)^2|\mathcal{F}_\infty^0] \leq \frac{C_3 g_i t_n}{n} \int_0^{(u-h)t_n} Z_0(s) ds. \tag{36}
$$

Summarizing, we have that

$$
\begin{aligned}
E[\Phi_{i,3}(u,h;n)|\mathcal{F}_\infty^0] &\leq \frac{C_1 t_n^2 g_i h}{n e^{t_n x_i(u-h)}} \int_0^{(u-h)t_n} Z_0(s) e^{-x_i s} ds + \frac{C_2 g_i t_n}{n} \int_{(u-h)t_n}^{ut_n} Z_0(s) e^{-x_i s} ds \\
&\quad + e^{2ht_n(\lambda_1 - x_i)} \left(ht_n(\lambda_1 - x_i) \right)^2 \frac{C_3 g_i t_n}{n} \int_0^{(u-h)t_n} Z_0(s) e^{-2x_i s} ds.
\end{aligned}
$$

Equation (36) implies that Claim A can be applied to the $\Psi_{j,3}$ term (with $\gamma_1 = 2$). It thus remains to establish an appropriate bound for $\Phi_{i,3}$, which requires showing that condition (A) applies to $t_n e^{-x_i t_n (u-h)}$ and $e^{2ht_n(\lambda_1 - x_i)}$. Since

$$t_n e^{-x_i t_n (u-h)} e^{2ut_n(\lambda_1 - x_i)} \le \left(t_n e^{-ax_i t} e^{ut_n(\lambda_1 - x_i)}\right) * e^{ut_n(x_i - \lambda_1)}$$

and the first term on the RHS of the inequality is bounded in i and n we see that we can apply condition (A). In addition, condition (A) applies to $e^{2ht_n(\lambda_1 - x_i)}$ due to the constraint $h \le u - a$ and that $\ell_3 = 2$.

For L_4 we have a single sum with

$$E[\Psi_{j,4}(u,h;n)] = e^{2ht_n(x_j - \lambda_1)} E\left[H_j(u,h;n)\left(M_j(ut_n) - e^{ht_n(\lambda_1 - x_j)} M_j\left((u-h)t_n\right)\right)^2\right].$$

If we define

$$X_j(u;n) = n^{\alpha-1} e^{-x_j u t_n} \hat{Z}_{1,j}(ut_n) + \frac{g_j Z_0(ut_n)}{n} \tag{37}$$

and then using the expressions for $\mathrm{Var}(\tilde{Z}_{1,j}(t))$ and $E[\tilde{Z}_{1,j}(t)^2]$ we can establish that there is $C > 0$ such that

$$H_j(u,h;n) \le Cht_n^2 e^{-x_j u t_n} X_j(u,n). \tag{38}$$

We can apply bound (38) to see that

$$E[\Psi_{j,4}(u,h;n)] \le Cht_n^2 e^{-x_j u t_n} e^{2ht_n(x_j - \lambda_1)} E\left[X_j(u;n)\left(M_j(ut_n) - e^{ht_n(\lambda_1 - x_j)} M_j\left((u-h)t_n\right)\right)^2\right]. \tag{39}$$

We now analyze the expected value in the preceding display

$$\begin{aligned}
&E\left[X_j(u;n)\left(M_j(ut_n) - e^{ht_n(\lambda_1 - x_j)} M_j\left((u-h)t_n\right)\right)^2\right], \\
&= E\left[X_j(u;n)\left(M_j(ut_n) - e^{ht_n(\lambda_1 - x_j)} M_j\left((u-h)t_n\right)\right)^2; X_j(u,n) \le 1/\sqrt{h}\right] \\
&\quad + E\left[X_j(u;n)\left(M_j(ut_n) - e^{ht_n(\lambda_1 - x_j)} M_j\left((u-h)t_n\right)\right)^2; X_j(u,n) > 1/\sqrt{h}\right] \\
&\le h^{-1/2} E\left[\left(M_j(ut_n) - e^{ht_n(\lambda_1 - x_j)} M_j\left((u-h)t_n\right)\right)^2\right] \\
&\quad + h^{1/2}\left(E\left[\left(M_j(ut_n) - e^{ht_n(\lambda_1 - x_j)} M_j\left((u-h)t_n\right)\right)^4\right] E\left[X_j(u;n)^4\right]\right)^{1/2}
\end{aligned}$$

where the inequality follows from two applications of the Cauchy-Schwarz inequality and one application of the Chebyshev inequality. We will now establish the following bounds

$$E\left[\left(M_j(ut_n) - e^{ht_n(\lambda_1 - x_j)} M_j\left((u-h)t_n\right)\right)^2\right] = O(g_j h t_n^3 e^{2ht_n(\lambda_1 - x_j)}), \tag{40}$$

and

$$E\left[\left(M_j(ut_n) - e^{ht_n(\lambda_1 - x_j)} M_j\left((u-h)t_n\right)\right)^4\right] E\left[X_j(u;n)^4\right] = O(g_j^2 t_n^3 e^{4ht_n(\lambda_1 - x_j)}). \tag{41}$$

Note that if we establish these bounds then (39) will imply that we have a function ρ_j of order $t_n^4 e^{-x_j u t_n}$. Since $k_4 = 2$, we see that condition (A) applies to this function, i.e., $t_n^4 e^{-x_j u t_n} e^{u t_n (x_j - \lambda_1)} = O(1)$. Using the martingale property we can see that

$$\begin{aligned}
&E\left[\left(M_j(ut_n) - e^{ht_n(\lambda_1 - x_j)} M_j\left((u-h)t_n\right)\right)^2\right] \\
&= E[H_j(u-h,h;n)] + \left(e^{ht_n(\lambda_1 - x_j)} - 1\right)^2 E\left[M_j\left((u-h)t_n\right)^2\right] \\
&\le Cht_n^2 e^{-x_j u t_n} E[X_j(u;n)] + \left(ht_n(\lambda_1 - x_j)\right)^2 e^{2ht_n(\lambda_1 - x_j)} E\left[M_j\left((u-h)t_n\right)^2\right].
\end{aligned}$$

The result in (40) now follows by observing that $E[X_j(u;n)] = O(g_j)$ and $E[M((u-h)t_n)^2] = O(g_j t_n)$. We can use the same 4th moment analysis as in the study of L_2 to conclude that

$$E\left[\left(M_j(ut_n) - e^{ht_n(\lambda_1 - x_j)} M_j\left((u-h)t_n\right)\right)^4\right] = O(g_j t_n e^{4ht_n(\lambda_1 - x_j)}),$$

and $E[X_j(u;n)^4] = O(g_j)$, thus establishing (41).

The last remaining term, L_5, which is a double sum with

$$\begin{aligned}
\Psi_{j,5}(u,h;n) &= \left(M_j(ut_n) - e^{ht_n(\lambda_1 - x_j)} M_j\left((u-h)t_n\right)\right)^2 \\
\Phi_{i,5}(u,h;n) &= e^{2ht_n(x_i - \lambda_1)} H_i(u,h;n).
\end{aligned}$$

We first consider $\Phi_{i,5}$

$$\begin{aligned}
&e^{ut_n(x_i - \lambda_1)} E[H_i(u,h;n)|\mathcal{F}_\infty^0] \le Cht_n^2 e^{-\lambda_1 u t_n} E[X_i(u;n)|\mathcal{F}_\infty^0] \\
&= \frac{Cht_n^2 e^{-\lambda_1 u t_n} g_i}{n} \left(\int_0^{ut_n} Z_0(s) e^{-x_i s} ds + Z_0(ut_n)\right),
\end{aligned}$$

to which we can clearly apply Claim A.

We next follow the analysis of the L_3 term to calculate that

$$\begin{aligned}
E[\Psi_{j,5}(u,h;n)|\mathcal{F}_\infty^0] &\le E[\tilde{H}_j(u-h,n)|\mathcal{F}_\infty^0] \\
&\quad + e^{2ht_n(\lambda_1 - x_j)} \left(ht_n(\lambda_1 - x_j)\right)^2 E\left[M_j((u-h)t_n)^2|\mathcal{F}_\infty^0\right] \\
&\le \frac{c_1 t_n^2 g_j h}{n e^{t_n x_i (u-h)}} \int_0^{(u-h)t_n} Z_0(s) e^{-x_j s} ds + \frac{c_2 g_j t_n}{n} \int_{(u-h)t_n}^{ut_n} Z_0(s) e^{-x_j s} ds \\
&\quad + e^{2ht_n(\lambda_1 - x_j)} \left(ht_n(\lambda_1 - x_j)\right)^2 \frac{c_3 g_j t_n}{n} \int_0^{(u-h)t_n} Z_0(s) e^{-2x_j s} ds.
\end{aligned}$$

The analysis of the first and third term follow exactly as in the study of the L_3 term. For the middle term we are missing a power of h, but this follows by considering the expected value of the product with $\Phi_{i,5}$. In particular, it is easily established that there exists constants $\delta_n \to 0$ such that

$$\frac{1}{n^2} E\left[\left(\int_0^{ut_n} Z_0(s)e^{-x_i s}ds + Z_0(ut_n)\right)\int_{(u-h)t_n}^{ut_n} Z_0(s)e^{-x_j s}ds\right]$$
$$= \frac{1}{n^2}\int_0^{ut_n}\int_{(u-h)t_n}^{ut_n} E[Z_0(s)Z_0(y)]e^{-x_i y - x_j s}dsdy + \frac{1}{n^2}\int_{(u-h)t_n}^{ut_n} E[Z_0(s)Z_0(ut_n)]e^{-x_j s}ds$$
$$= \delta_n h.$$

This result combined with the result of Claim A implies that there exists a positive constant $C > 0$ such that

$$L_5(u,h;n) = \sum_{i=1}^{\lambda_1(\triangle_n)} \sum_{j\neq i} e^{2ut_n(x_j-\lambda_1)} e^{2ut_n(x_i-\lambda_1)} E\left[\Phi_{i,5}(u,h;n)\Psi_{j,5}(u,h;n)\right]$$
$$\leq Ch^2.$$

□

4.2.5 Proof of Lemma 5

Proof. We prove the lower bound first. Recall that $e^{-\lambda t}B(t)$ is a non-negative martingale and thus $e^{-k\lambda t}B(t)^k$ is a submartingale. Therefore $E[B(t)^k] \geq e^{k\lambda t}$, which establishes the appropriate lower bound for the supercritical case. For the subcritical case recall that B is a non-negative integer valued process and therefore $E[B(t)^k] \geq E[B(t)] = e^{\lambda t}$.

To prove the upper bound first define the function $u(s) = a(f(s) - s)$, and the generating function for $s \in (0,1)$

$$F(s,t) = \sum_{n=0}^{\infty} s^n P(B(t) = n).$$

Next define

$$\ell_k(t) = \frac{\partial^k}{\partial s^k} F(s,t)\big|_{s=1} = \sum_{n=0}^{\infty} n(n-1)\cdots(n-k+1)P(B(t)=n). \tag{42}$$

Recall that $f^{(k)}(1) < \infty$ implies $E[B(t)^k] < \infty$ by Corollary III.6.2 of [2] and therefore $\ell_k(t) < E[B(t)^k] < \infty$.

Based on (42) we know that for each $k \geq 1$ there exists $C'_k > 0$ such that

$$\ell_k(t) \geq C'_k \sum_{n=k}^{\infty} n^k P(B(t) = n)$$
$$= C'_k E[B(t)^k] - C'_k E[B(t)^k; B(t) < k].$$

Since the branching process goes to infinity or extinct with probability 1 we know that $E[B(t)^k; B(t) < k] \to 0$ as $t \to \infty$, and in particular in the supercritical case we then have that

$$\lim_{t\to\infty} \frac{E[B(t)^k; B(t) \le k]}{E[B(t)^k]} = 0,$$

and thus there exists C_k such that $\ell_k(t) \ge C_k E[B(t)^k]$. In the subcritical case we can use Yaglom's theorem to see that $P(Z(t) > 0) \sim e^{\lambda t}/b$ for a positive constant b. Therefore there exist C_k and $\hat{C}_k$ such that

$$C_k E[B(t)^k] \le \ell_k(t) + \hat{C}_k e^{\lambda t}.$$

Thus, in order to prove the desired upper bound on $E[B(t)^k]$ it suffices to establish the upper bound on $\ell_k(t)$. In both the super and subcritical cases the proof for the upper bound on ℓ_k is carried out via induction on k, with the induction step proven via the forward equation

$$\frac{\partial}{\partial t} F(s,t) = u(s) \frac{\partial}{\partial s} F(s,t). \tag{43}$$

We first assume that $\lambda > 0$, and will prove that for non-negative integer k, $\ell_k(t) = O(e^{\lambda k t})$. Since $E[B(t)] = e^{\lambda t}$, the base case for the induction follows. Next assume that for $k > 1$ and $j \le k-1$, $\ell_j(t) = O(e^{\lambda j t})$. Then via (43) we have that

$$\begin{aligned}
\ell_k'(t) &= \sum_{j=0}^{k} \binom{k}{j} \ell_{k-j+1}(t) u^{(j)}(1) = \sum_{j=1}^{k} \binom{k}{j} \ell_{k-j+1}(t) u^{(j)}(1) \\
&= \lambda k \ell_k(t) + \sum_{j=2}^{k} \binom{k}{j} \ell_{k-j+1}(t) u^{(j)}(1),
\end{aligned}$$

where the first equality follows from $u(1) = 0$. Combining the previous display with the initial condition $\ell_k(0) = 0$ and then applying the induction hypothesis we have that there exists non-negative constants $\alpha_{k,j}$ such that

$$\begin{aligned}
e^{-\lambda k t} \ell_k(t) &= \sum_{j=2}^{k} u^{(j)}(1) \binom{k}{j} \int_0^t e^{-\lambda k s} \ell_{k-j+1}(s) ds \\
&\le \sum_{j=2}^{k} \alpha_{k,j} \int_0^t e^{-\lambda s(j-1)} ds = \sum_{j=2}^{k} \frac{\alpha_{k,j}}{\lambda(j-1)} \left(1 - e^{-\lambda t(j-1)}\right),
\end{aligned}$$

thus establishing the induction hypothesis and the desired result. The subcritical case is analyzed via the same methods. □

4.3 Proof of Proposition 1

We consider the quantity

$$A_2(n,y) \equiv \sup_{u\in[a,u_n^-(y)]} v(n,u)^{-1/2}\left(\phi_1(ut_n)-\phi_0(ut_n)\right),$$

which will be slightly different in Cases I and II. First consider Case I. Observe that due to monotonicity in u we have that

$$A_2(n,y) = v(n,u_n^-(y))^{-1/2}(\phi_1(u_n^-(y)t_n)-\phi_0(u_n^-(y)t_n)).$$

We can calculate that

$$A_2(n,y) = \frac{n^{(\alpha-1)/2-u_n^-(y)(\lambda_1/r+1)+1}}{\sqrt{\kappa_1}}\left(\frac{\mu}{n^\alpha(\lambda_1+r)}n^{(1+\lambda_1/r)u_n^-(y)}-1-\frac{\mu}{n^\alpha(\lambda_1+r)}\right).$$

Referring to the definition of u_n we see that

$$e^{u_n(\lambda_1+r)} = 1+\frac{n^\alpha(\lambda_1+r)}{\mu}$$

and therefore

$$\frac{\mu}{n^\alpha(\lambda_1+r)}n^{u_n^-(y)(1+\lambda_1/r)} = \left(\frac{\mu+n^\alpha(\lambda_1+r)}{n^\alpha(\lambda_1+r)}\right)e^{-y(\lambda_1+r)/s_n}. \tag{44}$$

Based on this we can rewrite $A_2(n,y)$ as follows

$$\frac{\mu n^{(1-\alpha)/2}e^{y(\lambda_1+r)/s_n}}{\sqrt{\kappa_1}(\lambda_1+r)}\left(e^{-y(\lambda_1+r)/s_n}-1\right).$$

We see that if we choose $s_n = n^{(1-\alpha)/2}$ then

$$A_2(n,y) \to -\frac{y\mu}{\sqrt{\kappa_1}}.$$

In case II the situation is a bit more complicated. We have that

$$\begin{aligned}
&v(n,u)^{-1/2}\left(\phi_1(ut_n)-\phi_0(ut_n)\right)\\
&= \left(\frac{un^{\alpha-1}\log n}{\kappa_2}\right)^{1/2} n^{-\lambda_1 u/r}\left[\mu n^{1-\alpha}\int_0^{\lambda_1}\int_0^{ut_n} g(x)e^{ut_nx-s(r+x)}dsdx - n^{1-u}\right]\\
&= \left(\frac{un^{\alpha-1}\log n}{\kappa_2}\right)^{1/2} n^{1-u(\lambda_1+r)/r}\left[\frac{\mu}{n^\alpha}\int_0^{\lambda_1}\frac{g(x)}{r+x}\left(e^{ut_n(r+x)}-1\right)dx-1\right]\\
&= \left(\frac{un^{\alpha-1}\log n}{\kappa_2}\right)^{1/2} n^{1-u(\lambda_1+r)/r}f_n(u).
\end{aligned}$$

We first establish that the function f_n has a unique root in $u_*(n) \in (0,1)$, and then approximate the root. First observe that, $f_n(0) = -1$ and that for sufficiently large n, $f_n(1) > 0$, the monotonicity of f_n establishes the uniqueness. A better localization of the root is obtained by considering

$$\begin{aligned} f_n\left(\frac{\alpha r}{\lambda_1 + r}\right) &= \frac{\mu}{n^\alpha}\int_0^{\lambda_1} \frac{g(x)}{r+x}\left(n^{\frac{\alpha(r+x)}{r+\lambda_1}} - 1\right) dx - 1 \\ &= \mu \int_0^{\lambda_1} \frac{g(x)}{r+x}\left(\exp\left[\alpha\left(\frac{x-\lambda_1}{r+\lambda_1}\right)\log n\right] - n^{-\alpha}\right) dx - 1 < 0, \end{aligned}$$

where the final inequality follows by applying the change of variable $z = t_n(\lambda_1 - x)$ to the first integral. This gives an improved lower bound on $u_*(n)$, and an improved upper bound is achieved by considering

$$\begin{aligned} &f_n\left(\frac{\alpha r}{\lambda_1 + r}\left(1 + \frac{3\log\log n}{2\alpha \log n}\right)\right) \\ &= \frac{\mu}{n^\alpha}\int_0^{\lambda_1} \frac{g(x)}{r+x}\left(\exp\left[\frac{\alpha(r+x)}{\lambda_1+r}\left(\log n + \frac{3\log\log n}{2\alpha}\right)\right] - 1\right) dx - 1 \\ &= \mu\left(\log n\right)^{3/2}\int_0^{\lambda_1} \frac{g(x)}{r+x}\left(\exp\left[\frac{\alpha(x-\lambda_1)}{\lambda_1+r}\left(\log n + \frac{3\log\log n}{2\alpha}\right)\right] - \frac{1}{n^\alpha(\log n)^{3/2}}\right) dx - 1. \end{aligned}$$

Then define $z_n = \log n + \frac{3}{2\alpha}\log\log n$ and use the change of measure $y = \alpha z_n(\lambda_1 - x)/(\lambda_1 + r)$ to see that

$$\begin{aligned} &f_n\left(\frac{\alpha r}{\lambda_1 + r}\left(1 + \frac{3\log\log n}{2\alpha \log n}\right)\right) \\ &= \frac{\mu(\log n)^{3/2}}{z_n}\int_0^{\frac{\alpha z_n}{\lambda_1 + r}} \frac{g\left(\lambda_1 - \frac{y(\lambda_1+r)}{\alpha z_n}\right)}{r + \lambda_1 - \frac{y(\lambda_1+r)}{z_n}} e^{-z} dz - 1 - \frac{\mu}{n^\alpha}\int_0^{\lambda_1}\frac{g(x)}{r+x}dx \\ &= \frac{2\alpha\mu\left(\log n\right)^{1/2}}{2\alpha + 3\log\log n/\log n}\left(\frac{g(\lambda_1)}{r+\lambda_1} + o(1)\right) - 1 - \frac{\mu}{n^\alpha}\int_0^{\lambda_1}\frac{g(x)}{r+x}dx, \end{aligned}$$

which is clearly positive for n sufficiently large. The final equality in the previous display follows from the dominated convergence theorem. We can now conclude that for n sufficiently large

$$u_*(n) \in \left(\frac{\alpha r}{\lambda_1 + r}, \frac{\alpha r}{\lambda_1 + r}\left(1 + \frac{3\log\log n}{2\alpha\log n}\right)\right). \tag{45}$$

Therefore $u_*(n) \to \frac{\alpha r}{\lambda_1 + r}$ as $n \to \infty$. We define

$$\psi(u) = u^{1/2} n^{-u(\lambda_1+r)/r}\left(\frac{\mu}{n^\alpha}\int_0^{\lambda_1}\frac{g(x)}{r+x}\left(e^{ut_n(r+x)} - 1\right) dx - 1\right),$$

and establish the claim

$$\arg\max_{u\in[a,u_n^+(y)]} \psi(u) = u_n^+(y). \tag{46}$$

This claim will of course follow by establishing that $\psi' > 0$ for $u \in [a, u_n^+(y)]$. Thus consider

$$\begin{aligned}
\psi'(u) = {} & \frac{\mu}{n^\alpha t_n}\int_0^{\lambda_1 t_n} \frac{g\left(\lambda_1 - y/t_n\right)e^{-uy}}{r+\lambda_1 - y/t_n}\left(\frac{1}{2\sqrt{u}} - y\sqrt{u}\right)dy + e^{-ut_n(\lambda_1+r)}\left(t_n\sqrt{u} - \frac{1}{2\sqrt{u}}\right) \\
& + \frac{\mu e^{-ut_n(\lambda_1+r)}}{n^\alpha}\left(t_n\sqrt{u} - \frac{1}{2\sqrt{u}}\right)\int_0^{\lambda_1}\frac{g(x)}{r+x}dx \\
\geq {} & \frac{-\mu\sqrt{u}}{n^\alpha t_n}\int_0^{\lambda_1 t_n} \frac{g\left(\lambda_1 - y/t_n\right)}{r+\lambda_1 - y/t_n} y e^{-uy}dy + e^{-ut_n(\lambda_1+r)}\left(t_n\sqrt{a} - \frac{1}{2\sqrt{a}}\right) \\
& + \frac{\mu e^{-ut_n(\lambda_1+r)}}{n^\alpha}\left(t_n\sqrt{a} - \frac{1}{2\sqrt{a}}\right)\int_0^{\lambda_1}\frac{g(x)}{r+x}dx
\end{aligned}$$

Using that $\max_{x\in[0,\lambda_1]} g(x) = G < \infty$ and $u \geq a$, we have that

$$\frac{\mu\sqrt{u}}{n^\alpha t_n}\int_0^{\lambda_1 t_n} \frac{g\left(\lambda_1 - y/t_n\right)}{r+\lambda_1 - y/t_n} y e^{-uy}dy \leq \frac{\mu\sqrt{u}G}{rn^\alpha t_n}\int_0^\infty y e^{-uy}dy \leq \frac{\mu G}{a^{3/2}n^\alpha \log n}.$$

Next we use that $u \leq \frac{\alpha r}{\lambda_1+r}(1+3\log\log n/(2\alpha\log n)) + y/s_n$ to see

$$e^{-ut_n(\lambda_1+r)}\left(t_n\sqrt{a} - \frac{1}{2\sqrt{a}}\right) \geq \left(\frac{\sqrt{a}}{rn^\alpha(\log n)^{1/2}} - \frac{n^{-\alpha}}{2\sqrt{a}(\log n)^{3/2}}\right)\exp\left[-\frac{y(\lambda_1+r)\log n}{s_n}\right].$$

for sufficiently large n. Thus for sufficiently large n, $\psi'(u) > 0$ for $u \in [a, u_n^+(y)]$. Define the function

$$h_n(u) = \int_0^{\lambda_1 u t_n} g\left(\lambda_1 - \frac{x}{ut_n}\right)e^{-x}dx,$$

and then observe that

$$f_n'(u) = \frac{\mu n^{u(\lambda_1+r)/r-\alpha}}{u}h_n(u). \tag{47}$$

Based on this and the monotonicity result for ψ we have that

$$
\begin{aligned}
&\sup_{u\in[a,u_n^-(y)]} \left(\frac{un^{\alpha-1}\log n}{\kappa_2}\right)^{1/2} n^{1-u(\lambda_1+r)/r} f_n(u) \\
&= \left(\frac{u_n^-(y)n^{\alpha-1}\log n}{\kappa_2}\right)^{1/2} n^{1-u_n^-(y)(\lambda_1+r)/r} f_n\left(u_n^-(y)\right) \\
&= -\sqrt{\frac{u_n^-(y)n^{\alpha+1}\log n}{\kappa_2}} n^{-u_n^-(y)(\lambda_1+r)/r} \int_{u_n^-(y)}^{u_*(n)} \frac{\mu n^{z(\lambda_1+r)/r}}{zn^\alpha} h_n(z)dz \\
&= -\mu e^{y(\lambda_1+r)/s_n} \sqrt{\frac{u_n^-(y)n^{1-\alpha}\log n}{\kappa_2}} \int_{u_n^-(y)}^{u_*(n)} \frac{e^{t_n(\lambda_1+r)(z-u_*(n))}}{z} h_n(z)dz \\
&= -\mu r e^{y(\lambda_1+r)/s_n} \sqrt{\frac{u_n^-(y)n^{1-\alpha}}{\kappa_2\log n}} \int_0^{y/s_n} \frac{e^{-x(\lambda_1+r)} h_n\left(u_*(n)-\frac{x}{t_n}\right)}{u_*(n)-\frac{x}{t_n}} dx
\end{aligned}
\tag{48}
$$

Next note that

$$
\begin{aligned}
&g(\lambda_1) - \int_0^{\lambda_1 ut_n} e^{-x} g\left(\lambda_1 - \frac{x}{ut_n}\right) dx \\
&= g(\lambda_1)e^{-\lambda_1 ut_n} + \int_0^{\lambda_1 ut_n} e^{-x}\left(g(\lambda_1) - g\left(\lambda_1 - \frac{x}{ut_n}\right)\right) dx \\
&\leq g(\lambda_1)e^{-\lambda_1 at_n} + \int_0^{\lambda_1 t_n} e^{-ux}\left(g(\lambda_1) - g\left(\lambda_1 - \frac{x}{t_n}\right)\right) dx \\
&\leq g(\lambda_1)e^{-\lambda_1 at_n} + \int_0^{\lambda_1 t_n} e^{-ax}\left(g(\lambda_1) - g\left(\lambda_1 - \frac{x}{t_n}\right)\right) dx,
\end{aligned}
$$

and therefore $h_n(z) \to g(\lambda_1)$ uniformly for $z \in [a, 1]$. We can thus apply the fundamental theorem of calculus and the bounds in (45) to (48) to conclude that

$$
\lim_{n\to\infty} \sup_{u\in[a,u_n^-(y)]} \left(\frac{un^{\alpha-1}\log n}{\kappa_2}\right)^{1/2} n^{1-u(\lambda_1+r)/r} f_n(u) = -\mu y g(\lambda_1)\sqrt{\frac{r(\lambda_1+r)}{\alpha\kappa_2}}.
$$

4.4 Proof of Proposition 2

We first consider the term

$$
C_3(n,y) = \sup_{u\in[a,u_n^-(y)]} n^{(\alpha-1)/2-\lambda_1 u/r}\left(Z_0(ut_n) - \phi_0(ut_n)\right),
$$

which satisfies $C_3(n,y) = \Theta(A_3(n,y))$ in Case I, and is off by a factor of $\sqrt{\log n}$ in Case II. We claim that $C_3(n,y)$ converges to 0 as $n \to \infty$ if a is chosen appropriately.

Consider the square

$$\begin{aligned}(C_3(n,y))^2 &\le n^{\alpha-1} \sup_{u\in[a,u_n^-(y)]} n^{-2\lambda_1 u/r}\left(Z_0(ut_n)-\phi_0(ut_n)\right)^2\\ &= n^{\alpha-1} \sup_{u\in[a,u_n^-(y)]} n^{2u}\left(Z_0(ut_n)-\phi_0(ut_n)\right)^2 n^{-2u(\lambda_1/r+1)}\\ &\le n^{\alpha-1-2a(\lambda_1/r+1)} \sup_{u\in[a,u_n^-(y)]} n^{2u}\left(Z_0(ut_n)-\phi_0(ut_n)\right)^2.\end{aligned}$$

Now observe that $n^{2u}\left(Z_0(ut_n)-\phi_0(ut_n)\right)^2$ is a submartingale in the parameter u and therefore for $\varepsilon>0$

$$\mathbb{P}\left(n^{\alpha-1-2a(\lambda_1/r+1)} \sup_{u\in[a,u_n^-(y)]} n^{2u}\left(Z_0(ut_n)-\phi_0(ut_n)\right)^2>\varepsilon\right) \tag{49}$$

$$\le \frac{n^{\alpha-1-2a(\lambda_1/r+1)}}{\varepsilon} n^{2u_n^-(y)}\mathrm{Var}(Z_0(u_n^-(y)t_n)) = O\left(n^{u_n^-(y)+\alpha-2a(\lambda_1/r+1)}\right), \tag{50}$$

where the final equality follows from the result $\mathrm{Var}Z_0(ut_n)=O(n^{1-u})$. From the definition of $u_n^-(y)$ we know that in the setting $G(dx)=\delta_{\lambda_1}$

$$\begin{aligned}&u_n^-(y)+\alpha-2a(1+\lambda_1/r)\\ &= \frac{r\log\left(1+n^\alpha(\lambda_1+r)/\mu\right)}{(\lambda_1+r)\log n}-\frac{y}{s_n}+\alpha-2a(1+\lambda_1/r)\\ &\le \frac{r\log\left(cn^\alpha(\lambda_1+r)\right)}{(\lambda_1+r)\log n}+\alpha-2a(1+\lambda_1/r)\\ &= \alpha\left(\frac{\lambda_1+2r}{\lambda_1+r}\right)-2a(1+\lambda_1/r)+\frac{cr\log(\lambda+r)}{(\lambda_1+r)\log n}.\end{aligned} \tag{51}$$

where $c>(\lambda_1+r)/\mu+1$. In the setting $G(dx)=g(x)dx$ we just have an extra term of the form $\log\log n/\log n$ and thus the result holds in this case as well. Based on (49), showing that $A_3(n)\to 0$ in probability we must show that the expression on the RHS of the previous display is bounded below zero for n sufficiently large. This requires that

$$\alpha\left(\frac{\lambda_1+2r}{\lambda_1+r}\right)<\frac{2a(\lambda_1+r)}{r}$$

or after rearranging terms

$$a>\frac{\alpha r(\lambda_1+2r)}{2(\lambda_1+r)^2}=\left(\frac{\alpha r}{\lambda_1+r}\right)\left(\frac{\lambda_1+2r}{2(\lambda_1+r)}\right).$$

Since

$$\left(\frac{\alpha r}{\lambda_1+r}\right)\left(\frac{\lambda_1+2r}{2(\lambda_1+r)}\right)<u_n^-(y)$$

for sufficiently large n the result follows by choosing $a\in[\left(\frac{\alpha r}{\lambda_1+r}\right)\left(\frac{\lambda_1+2r}{2(\lambda_1+r)}\right),u_n^-(y)]$. Note that this also implies that $\sqrt{\log n}C_3(n,y)\to 0$ as $n\to\infty$. Thus the term A_3 goes to 0 in both Cases I and II.

5 Acknowledgements

J.F is partially supported by the McKnight Foundation, a National Cancer Institute Physical Sciences Oncology Center Transnetwork Grant, and the National Science Foundation (NSF DMS-1224362 and NSF DMS-1349724). K.L is partially supported by National Science Foundation (NSF DMS-1224362 and NSF CMMI-1362236). J.Z is partially supported by the National Science Foundation (NSF DMS-12-24362).

References

[1] B. Aldridge, M. Fernandez-Suarez, and et. al. Asymmetry and aging of mycobacterial cells lead to variable growth and antibiotic susceptibility. *Science*, p. 100–104, 2011.

[2] K.B. Athreya and P.B. Ney. *Branching processes.* Dover, 2004.

[3] P. Billingsley. *Convergence of Probability Measures Second Edition.* John Wiley and Sons, 1999.

[4] J. Chmielecki, J. Foo, and et al. Optimization of dosing for EGFR-mutant non-small cell lung cancer with evolutionary cancer modeling. *Science Translational Medicine*, 3:90ra59, 2011.

[5] R. Durrett, J. Foo, K. Leder, J. Mayberry, and F. Michor. Evolutionary dynamics of tumor progression with random fitness values. *Theoretical Population Biology*, 78:54–66, 2010.

[6] J. Foo and K. Leder. Dynamics of cancer recurrence. *Annals of Applied Probability*, 23:1437–1468, 2013.

[7] P. Jagers, F. Klebaner, and S. Sagitov. Markovian paths to extinction. *Advances in Applied Probability*, 39:569–587, 2007.

[8] P. Jagers, F. Klebaner, and S. Sagitov. On the path to extinction. *Proceedings of the National Academy of Sciences*, 104:6107–6111, 2007.

[9] N.L. Komarova and D. Wodarz. Drug resistance in cancer: Principles of emergence and prevention. *PNAS*, 102:9714–9719, 2005.

[10] K. Leder, J. Foo, and et. al. Diversity in pre-existing resistance to BCR-ABL inhibitors in chronic myeloid leukemia. *PLoS ONE*, e27682, 2011.

[11] R.M. Loynes. A criterion for tightness for a sequence of martingales. *Annals of Probability*, 4:859–862, 1976.

[12] D. Tyson and et. al. Fractional proliferation: A method to deconvolve cell population dynamics from single-cell data. *Nature Methods*, pages 923–928, 2012.