Submitted to the Annals of Applied Probability
arXiv: arXiv:0000.0000

DYNAMICS OF CANCER RECURRENCE

By Jasmine Foo*, and Kevin Leder*

*University of Minnesota**

Mutation-induced drug resistance in cancer often causes the failure of therapies and cancer recurrence, despite an initial tumor reduction. The timing of such cancer recurrence is governed by a balance between several factors such as initial tumor size, mutation rates, and growth kinetics of drug-sensitive and resistance cells. To study this phenomenon we characterize the dynamics of escape from extinction of a subcritical branching process, where the establishment of a clone of escape mutants can lead to total population growth after the initial decline. We derive uniform in time approximations for the paths of the escape process and its components, in the limit as the initial population size tends to infinity and the mutation rate tends to zero. In addition, two stochastic times important in cancer recurrence will be characterized: (i) the time at which the total population size first begins to rebound (i.e. become supercritical) during treatment, and (ii) the first time at which the resistant cell population begins to dominate the tumor.

1. Introduction. We consider a situation arising from population genetics, where a population with net negative growth rate can escape certain extinction via creation of a new mutant type. This scenario arises in a variety of biological and medical applications. In particular, we consider the following scenario in which a population of drug-sensitive cancer cells is placed under therapy, leading to a sustained overall decline in tumor size. However, drug-resistance mutations may arise in the population, conferring a net positive growth rate to mutated cells and their progeny under therapy. If a mutant arises prior to extinction of the original population and forms a viable, growing subpopulation, then the population has 'escaped' extinction. These types of escape events due to acquired resistance cause the failure of many drugs including antibiotics, cancer therapies and anti-viral therapies. In the cancer setting, the discovery of new molecularly targeted therapies has lead to dramatic successes in tumor reduction in the past decade; however the majority of these therapies fail due to the development of drug resistance and subsequent increase in tumor burden and progression of disease. Examples of targeted therapies for which acquired resistance exists

AMS 2000 subject classifications: Primary 60J80
Keywords and phrases: branching processes, population genetics, cancer

include erlotinib/gefitinib in EGFR-mutant non-small cell lung cancer, imatinib, dasatinib, or nilotinib in BCR-ABL driven chronic myeloid leukemia, and vemurafenib in BRAF-mutant melanoma.

There has been a significant amount of previous work in the cancer modeling literature on understanding the evolutionary dynamics of drug resistance in cancer. For example, using stochastic processes with a differentiation hierarchy to represent the sensitive and resistant cells of a tumor, Coldman and Goldie studied the emergence of resistance to one or two drugs [3, 5, 6]. In a different twist, Harnevo and Agur studied drug resistance emerging due to oncogene amplification using a stochastic branching process model [7, 8]. Others have used multi-type branching process models to study the probability of resistance emerging due to point mutations in a variety of situations, e.g. [15, 9]. Komarova and Wodarz also utilized a multi-type branching model to investigate the general situation in which k mutations are required to confer resistance against k drugs [12, 13]. Most recently, in [4] the authors considered an inhomogeneous process wherein the birth and death rates of both sensitive and resistant cells are dependent upon a temporally varying drug concentration profile, to accommodate the effects of pharmacokinetic dynamics as the drug is metabolized over time. The analysis in most of these works has been focused on calculations of the eventual probability of developing resistance and the resistant population size, rather than the variable timing of tumor recurrence.

In addition to work specifically related to mathematical modeling of cancer recurrence, we also discuss some mathematical contributions to the study of extinction paths in subcritical branching processes and the dynamics of escape in this context. In particular, in [10] Jagers and co-authors considered large population approximations of 'the path to extinction' in Markovian sub-critical branching processes. In this work they established convergence of finite dimensional distributions of these paths viewed on the time scale of extinction. The follow up work [11] generalized these results to a broader class of inter-arrival times (i.e., distributions more general than exponential). Sehl and et. al. investigated the limiting moments of extinction times of subcritical branching processes, and used this as a tool for investigating the effects of various cancer therapies on healthy tissue [20]. Lastly, Sagitov and Serra characterized the asymptotic structure for BGW process with escape, as mutation rate $\mu \to 0$, conditioned on successful escape, which is an important asymptotic regime in many problems such as the evolution of new species [19].

A typical solid tumor has a density between 10^7 and 10^9 cancer cells per cubic centimeter, [14]. Therefore, in this work we are interested in deriving

path approximations of the escape process that are uniform in time, in the regime of a very large initial population. In the large population limit, it is tempting to assume that the stochastic model can be approximated by a purely deterministic model. However, a simple comparison of the mean behavior of the stochastic model with a deterministic model illustrates that it is important to consider the stochasticity of the extinction time. Here, we develop limiting stochastic approximations for the population process that greatly simplify the population process model while maintaining the stochastic extinction time behavior. Interesting earlier work by Jagers, Sagitov, et. al. established convergence of the finite dimensional distributions for the declining sensitive cell populations on the time scale of extinction, leaving open the question of tightness [10]. In the present work we first construct nearly-deterministic uniform in time limit approximations to the both the declining sensitive population paths as well as the supercritical resistant cell escape paths. Then, tightness of the joint sensitive and resistant process can be established as a simple consequence of these approximations, yielding the weak convergence result in simpler fashion than via direct analysis of the joint process.

We then use these approximations to characterize the distribution of 'turnaround times,' at which the total population size switches from subcritical to supercritical. In the clinical context, this represents the time at which progression of disease is observed through serial tumor scans or bloodwork (in leukemias); thus the ability to characterize and predict this time is of significant prognostic interest. In addition, we characterize the 'crossover time' at which the resistant mutants first overtake the original type in the population. Estimates of crossover times, and more generally the times at which certain composition thresholds are reached, are extremely useful in clinical decision-making. For example, when simultaneous combination therapies are considered, understanding these random times allows for informed decisions on the optimal time to switch to another therapy and thus 'target' a different subpopulation of cells within the tumor. Figure 1 illustrates these times in a sample path simulation of the process, in addition to a sample distribution of turnaround times. Our results are derived in the framework where the time scale of the processes is sped up by the extinction time of the original population, a natural time scale since this time represents the maximum length of effectiveness of the drug. We restrict our attention to binary branching processes which are appropriate for modeling cancer cell populations undergoing binary division; however, these results can be extended to study more general offspring distributions, and thus may be useful for studying escape dynamics in viral populations, for instance.

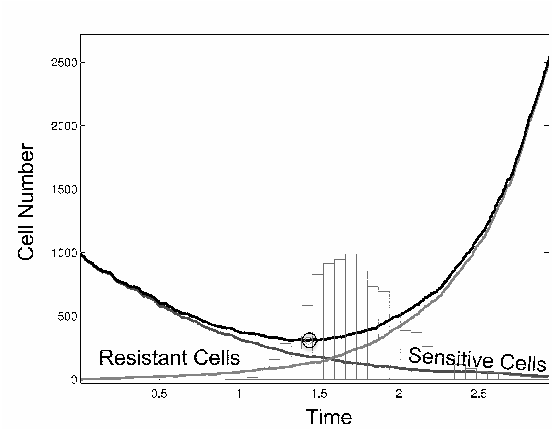

FIG 1. *Sample simulation of escape dynamics (population size vs time). Dark black line: total population size of the tumor, labeled grey lines: resistant and sensitive cell population size. The circle marks the minimum of the total tumor size process (i.e. the turnaround time), and the point at which red and blue lines cross is the crossover time. A histogram plotting the distribution of the minimum turnaround time for each sample path is plotted in green in the background. Parameters: starting population 1000 sensitive, 0 resistant. Net growth rate of sensitive and resistant birth-death processes are -1.0 and 2.0, respectively, and the mutation rate* $\mu = 0.01$.

The rest of the paper is organized as follows. In section 2 we introduce the model and discuss earlier results in the field. In section 3 we present some results on the mean of the resistant cell population at multiples of the extinction time. In section 4 we present a path approximation result where we show that the limit process uniformly approximates both the sensitive and resistant cell process on the time scale of the extinction time of the sensitive cells. We determine limiting distributions of the crossover time when the resistant cell population first becomes dominant, and the random time of disease progression or the 'turnaround' time. In section 5 we briefly illustrate an application of these results to studying the the time of disease recurrence due to drug resistance in non-small cell lung cancer (NSCLC). In section 7 we present the proofs of our main results.

Throughout the paper we use the following standard Landau asymptotic notation for non-negative functions $f(\cdot)$ and $g(\cdot)$: $f(x) = O(g(x))$ means that $f(x) \leq cg(x)$ for some $c \in (0, \infty)$, $f(x) = \Omega(g(x))$ if and only if $f(x) \geq cg(x)$, $f(x) = o(g(x))$ holds if and only if $f(x)/g(x) \to 0$ as $x \to \infty$, and lastly $f(x) \sim g(x)$ holds if and only if $f(x)/g(x) \to 1$ as $x \to \infty$.

2. Model and Previous Work. In this section we introduce the mathematical model and notations, and review previous results on related problems. We start with an initial population of drug sensitive cells with size x. This population $Z_0(t)$ is modeled as a subcritical Markovian binary branching process which declines during treatment with net growth rate $\lambda_0 < 0$, birth rate r_0 and death rate d_0, we will also use the notation $|\lambda_0| = r$. Resistance mutations arise at rate $\mu_x Z_0(t)$, and each of these mutations give rise to a supercritical Markovian binary branching process initialized by one mutant cell with net growth rate $\lambda_1 > 0$. We set $\mu_x = \mu x^{-\alpha}$ for $\mu > 0$ and $\alpha \in (0, 1)$. The total population of mutants, which we will call 'resistant cells,' is denoted $Z_1(t)$. These processes are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In addition define the filtration $\mathcal{F}_t^i$ generated by $Z_j(s)$, for $s \leq t$ and $j \leq i$. Note that in this work, unless otherwise stated the expectation and probability operators are conditioned on the initial conditions $Z_0(0) = x$ and $Z_1(0) = 0$.

Since the net growth rate of the original population is negative it will go extinct eventually with probability 1. We will denote this time of extinction by T_x, where x denotes the starting population. The following limit theorem from [17] will prove useful throughout the rest of the paper

$$T_x - \frac{1}{r}\log x \Rightarrow \frac{1}{r}\left(\eta + \log c\right), \text{ as } x \to \infty \tag{2.1}$$

where η is a standard Gumbel random variable and c is the Yaglom constant for Z_0. For a binary branching process, the Yaglom constant has the form $(d_0 - r_0)/r_0$.

Previously, Jagers and colleagues [10] studied the paths to extinction in a subcritical Markovian branching process, which we will also call Z_0 starting at size x. They considered the process Z_0 on the time scale of the extinction time and established convergence in finite dimensional distributions as $x \to \infty$.

Theorem 1 (Jagers et al, 2007). *For $u \in [0, 1)$,*

$$x^{u-1} Z_0(uT_x) \xrightarrow{FD} c^{-u} e^{-u\eta}.$$

Similar results on convergence in finite dimensional distribution of subcritical branching processes with more general inter-arrival times were also shown in [11]. In addition, Kimmel and Wu generalized these results to consider the case of critical branching processes [22].

3. Mean of $Z_1(uT_x)$. In this section we examine the growth rate of the mean of Z_1. In addition we examine a common modeling assumption and note the importance of considering the tails of the extinction time T_x in studies of escape dynamics. We will first consider the expected resistant population at vT_x for some $v > 0$, (and temporarily assume $\alpha = 0$)

$$\mathbb{E}\left[Z_1(vT_x)\right] = \mathbb{E}\left[\mu T_x \int_0^{v\wedge 1} Z_0(uT_x)\exp\left(\lambda_1 T_x(v-u)\right)du\right]. \tag{3.1}$$

If we assume that sensitive cells follow a deterministic decay $Z_0(t) = xe^{\lambda_0 t}$ and approximate their extinction time as $T_x \approx -\frac{1}{\lambda_0}\log x$, then we can heuristicallylly estimate the expected value as

$$\begin{aligned}
\mathbb{E}[Z_1(vT_x)] &= \frac{\mu}{r}\log x \int_0^{v\wedge 1} x^{1-u} x^{\frac{\lambda_1}{r}(v-u)} du \\
&= \frac{\mu}{r} x^{1-\frac{\lambda_1}{\lambda_0}v} \log x \int_0^{v\wedge 1} x^{-u(1+\frac{\lambda_1}{r})} du \\
&= \frac{\mu}{\lambda_1-\lambda_0} x^{1+\frac{\lambda_1}{r}v}\left(1-\exp\left[-(v\wedge 1)(1+\frac{\lambda_1}{r})\log x\right]\right).
\end{aligned}$$

Thus we observe that this expected value is finite for all $v > 0$.

However, suppose that there is just a single sensitive cell at time $t = 0$ whose birth rate is $r_0 = 0$, and death rate r . Then of course the extinction time satisfies $T_1 \sim \exp(r)$ and by conditioning on this time we have

$$\mathbb{E}[Z_1(vT_x)] = \mu \int_0^\infty \mathbb{E}[e^{\lambda_1(vT_1-s)} 1_{\{T_1 \geq v/s\}}]ds.$$

Due to the exponential tails of T_1 we see that the above integral diverges to ∞ for $\lambda_1 v \geq r$ and we clearly see the importance of the randomness in the extinction time. The previous result easily applies to models with births and deaths in the sensitive cell population. In particular,

Proposition 1. *Let $v > 0$ and $\frac{\lambda_1 v}{r} > 1$ then for all x $\mathbb{E}[Z_1(vT_x)]$ is infinite.*

Proof. By conditioning on $Z_0(s), s > 0$ and then applying a change of measure we can write the integral of interest as

$$\begin{aligned}
\mathbb{E}\left[Z_1(vT_x)\right] &= \mathbb{E}\left[\mu T_x \int_0^{v\wedge 1} Z_0(uT_x)\exp\left[\lambda_1 T_x(v-u)\right]du\right] \\
&= \mu \int_0^{v\wedge 1}\int_0^\infty te^{\lambda_1 t(v-u)}\mathbb{E}\left[Z_0(uT_x) \mid T_x \in dt\right] g_x(t)dtdu,
\end{aligned}$$

where $g_x(t)dt = \mathbb{P}(T_x \in dt)$.

Noting the fact that if $u < 1$ then $E\left[Z_0(uT_x) \mid T_x \in dt\right] \geq 1$, we can bound this from below by

$$\begin{aligned}
\mathbb{E}\left[Z_1(vT_x)\right] &\geq \mu \int_0^{v\wedge 1} \int_0^\infty te^{\lambda_1 t(v-u)} g_x(t)dtdu \\
&= \mu \int_0^{v\wedge 1} \left(\mathbb{E}[T_x] + \lambda_1(v-u)\int_0^\infty e^{\lambda_1 s(v-u)} \int_s^\infty tg_x(t)dtds\right) du \\
&\geq \mu \int_0^{v\wedge 1} \left(\mathbb{E}[T_x] + \lambda_1(v-u)\int_0^\infty e^{\lambda_1 s(v-u)} s\int_s^\infty g_x(t)dtds\right) du \\
&\geq \mu \int_0^{v\wedge 1} \left(\mathbb{E}[T_x] + c\lambda_1(v-u)\int_{t_0}^\infty e^{\lambda_1 s(v-u)} se^{-rs}ds\right) du.
\end{aligned}$$

The final inequality is based on the fact that for $x \geq 1$, $\mathbb{P}(T_x > s) \geq \mathbb{P}(T_1 > s)$ and the asymptotic result that as $t \to \infty$, $\mathbb{P}(T_1 > t) \sim ce^{-rt}$. Considering the final equation in the previous display, we see that if $\lambda_1 v > r$ then for u sufficiently small the inner integral diverges to ∞. □

We can easily find the asymptotic growth rate of $\mathbb{E}[Z_1(vT_x)]$ as $x \to \infty$. Based on the previous subsection, we know that this is only meaningful if we consider $v \leq -\lambda_0/\lambda_1$, for simplicity we will just assume that $v \leq 1$ and $r = |\lambda_0| \geq \lambda_1$. Earlier heuristic calculations (where we set $\alpha = 0$) indicate that the mean of $Z_1(vT_x)$ grows like $x^{1+v\lambda_1/r}$ as $x \nearrow \infty$. In particular we have the following,

THEOREM 2. *Assume that $r \geq \lambda_1$, then for $v \in (0,1]$ and $\alpha \in (0,1)$ we have that*

$$\mathbb{E}\left[Z_1(vT_x)\right] \sim x^{1+v\lambda_1/r-\alpha}\frac{c^{\lambda_1 v/r}\mu\Gamma(1-\frac{\lambda_1 v}{r})}{\lambda_1 + r}.$$

We defer the proof of this result to a later section.

4. Paths of escape. We now establish an approximation theorem for the paths of the joint process $(Z_0(uT_x), Z_1(uT_x))$. In the large x limit, scaled versions of these paths can be approximated uniformly in time by a simple stochastic process whose only source of randomness arises from the stochasticity of the limit theorem for the extinction time.

Before beginning, we first establish some notation. We will work with scaled versions of the sensitive and resistant populations sped up in time.

Let us define $s_x(t) = \frac{1}{r}\log x + t$. For $u \in [0,1]$ and $t \in \mathbb{R}$, define

$$Z_0^x\left(us_x(t)\right) = x^{u-1} Z_0\left(u\left(\frac{1}{r}\log x + t\right)\right) \tag{4.1}$$
$$Z_1^x\left(us_x(t)\right) = x^{-\lambda_1 u/r - 1 + \alpha} Z_1\left(u\left(\frac{1}{r}\log x + t\right)\right).$$

Throughout the rest of the paper, the superscript x will denote scaling by the appropriate function of x. For ease of notation we introduce the following notation.

$$\phi_0^x(u,t) = \mathbb{E}Z_0^x\left(us_x(t)\right) = e^{\lambda_0 ut}$$
$$\phi_1^x(u,t) = \mathbb{E}Z_1^x\left(us_x(t)\right) = \frac{\mu e^{\lambda_1 ut}}{\lambda_1 - \lambda_0}\left(1 - e^{(\lambda_0-\lambda_1)ut} x^{\frac{(\lambda_0-\lambda_1)u}{r}}\right).$$

In addition we will sometimes need to work with the population processes sped up in time but not scaled in space, which are defined for $Z_i(us_x(t))$, for $i = 0, 1$ and their means are denoted by

$$\phi_i(u,t) = \mathbb{E}Z_i(us_x(t)).$$

In the following, we establish the approximation result by first showing that for any $t \in \mathbb{R}$ we can approximate the scaled joint process by its mean uniformly in u. This is done by martingale arguments and showing relevant second moments are uniformly bounded in x. We then prove that this approximation is uniform for t in compact sets, and that one can approximate $(Z_0(uT_x), Z_1(uT_x))$ uniformly in time by $(\phi_0(u, T_x - \frac{1}{r}\log x), \phi_1(u, T_x - \frac{1}{r}\log x))$, where the previous formula is interpreted as the mean functions ϕ_i^x evaluated at the random parameter $T_x - \frac{1}{r}\log x$. We begin with a result on the moments of Z_0 and Z_1.

Lemma 1. *Let $\tilde{Z}_1$ be a binary branching process starting from size one with birth rate r_1 and death rate d_1, then for $0 < s < s_x(t)$,*

(i)

$$\mathbb{E}[Z_1\left(s_x(t)\right)^2] = \frac{\mu^2}{x^{2\alpha}}\int_0^{s_x(t)}\int_0^{s_x(t)} \mathbb{E}[Z_0(s)Z_0(y)]e^{\lambda_1(s_x(t)-s)}e^{\lambda_1(s_x(t)-y)}dsdy$$
$$+ \frac{\mu}{x^\alpha}\int_0^{s_x(t)} \mathbb{E}Z_0(s)\mathbb{E}[\tilde{Z}_1(s_x(t)-s)^2]ds.$$

(ii)
$$\mathbb{E}[Z_0(s)Z_1\left(s_x(t)\right)] = \frac{\mu}{x^\alpha}\int_0^{s_x(t)} \mathbb{E}[Z_0(y)Z_0(s)]e^{\lambda_1(s_x(t)-y)}dy$$

(iii)

$$Var[Z_1(s_x(t))] = \frac{\mu^2}{x^{2\alpha}} \int_0^{s_x(t)} \int_0^{s_x(t)} Cov(Z_0(s), Z_0(y)) e^{\lambda_1(2s_x(t)-(s+y))} ds dy + \frac{\mu}{x^\alpha} \int_0^{s_x(t)} \mathbb{E}Z_0(s)\mathbb{E}[\tilde{Z}_1(s_x(t)-s)^2] ds.$$

The proof of this result can be found in section 7.

Lemma 1 allows us to establish the following result via the Doob's maximal inequality.

LEMMA 2. *For $a \in (0,1)$, $\varepsilon > 0$, and $t \in \mathbb{R}$,*

(i)

$$\lim_{x\to\infty} \mathbb{P}\left(\sup_{u\in[0,a]} |Z_0^x(us_x(t)) - \phi_0^x(u,t)| > \varepsilon \right) = 0$$

(ii)

$$\lim_{x\to\infty} \mathbb{P}\left(\sup_{u\in[0,1]} |Z_1^x(us_x(t)) - \phi_1^x(u,t)| > \varepsilon \right) = 0$$

The proof of this result can be found in section 7.

We can strengthen Lemma 2 by showing the convergence above is in fact uniform for t in a compact set.

LEMMA 3. *For $a \in (0,1)$, $\varepsilon > 0$, and $M > 0$,*

(i)

$$\lim_{x\to\infty} \mathbb{P}\left(\sup_{t\in[-M,M]} \sup_{u\in[0,a]} |Z_0^x(us_x(t)) - \phi_0^x(u,t)| > \varepsilon \right) = 0$$

(ii)

$$\lim_{x\to\infty} \mathbb{P}\left(\sup_{t\in[-M,M]} \sup_{u\in[0,1]} |Z_1^x(us_x(t)) - \phi_1^x(u,t)| > \varepsilon \right) = 0.$$

The result is established by showing that the probabilities in the statement of Lemma 2 are monotone in the parameter t. Again we defer the full proof until section 7.

Using this uniform approximation result, we establish the following theorem for the process paths evaluated at multiples of the Z_0 extinction time.

THEOREM 3. *For $a < 1$, $\varepsilon > 0$, and $\mu_x = \mu x^{-\alpha}$ where $\alpha \in (0,1)$,*

(i)

$$\lim_{x\to\infty} \mathbb{P}\left(\sup_{u\in[0,a]} x^{u-1}\left|Z_0(uT_x) - \phi_0\left(u, T_x - \frac{1}{r}\log x\right)\right| > \varepsilon\right) = 0,$$

(ii)

$$\lim_{x\to\infty} \mathbb{P}\left(\sup_{u\in[0,1]} x^{\alpha-u\lambda_1/r-1}\left|Z_1(uT_x) - \phi_1\left(u, T_x - \frac{1}{r}\log x\right)\right| > \varepsilon\right) = 0.$$

Proof. This result now follows directly from Lemma 3 and the result in (2.1). □

We define the following stochastic processes: if $u \in [0,1]$,

$$\psi_0(u) = e^{-u(\eta+\log c)}$$
$$\psi_1(u) = \begin{cases} \frac{\mu}{\lambda_1+r} e^{\frac{\lambda_1 u}{r}(\eta+\log c)}, & u > 0 \\ 0 & u = 0. \end{cases}$$

These processes represent the limits of our scaled population processes. However, note that ψ_1 is not right-continuous at 0 and therefore it is not possible to establish that the scaled population processes are tight on an interval of the form $[0,b]$ in the standard Skorokhod topology. This is a result of the massive influx of mutations near $t = 0$ in the unscaled process.

Next, we utilize Theorem 3 to establish weak convergence in the Skorokhod sense of the joint process.

Corollary 4. *For $\alpha \in [0,1)$ and $0 < a < b < 1$ the joint process*

$$\{(x^{u-1}Z_0(uT_x), x^{\alpha-\lambda_1 u/r-1}Z_1(uT_x)), u \in [a,b]\} \Rightarrow \{(\psi_0(u), \psi_1(u)), u \in [a,b]\}$$

as $x \to \infty$ in the standard Skorokhod topology, $D([a,b])$.

Proof. For ease of notation throughout this proof we will use the following notation

$$\boldsymbol{\phi}^x(u) = \left(\phi_0^x(u, T_x - \frac{1}{r}\log x), \phi_1^x(u, T_x - \frac{1}{r}\log x)\right).$$

Clearly from the result in Theorem 3 it suffices to prove that as $x \to \infty$

$$\left(\phi_0^x(\cdot, T_x - \frac{1}{r}\log x), \phi_1^x(\cdot, T_x - \frac{1}{r}\log x)\right) \Rightarrow (\psi_0(\cdot), \psi(\cdot))$$

in $D([a,b])$. We will carry this out via Theorem 13.3 of [1]. First, we observe that convergence in finite dimensional distributions follows from (2.1) and the continuous mapping theorem. It thus only remains to establish tightness. Since our limit functions are continuous at $u = a$ and $u = b$, it suffices to establish that for every $\varepsilon > 0$

$$\lim_{\delta\to 0}\limsup_{x\to\infty}\mathbb{P}(\omega_x^{''}(\delta) \geq \varepsilon) = 0, \tag{4.2}$$

where

$$\omega_x^{''}(\delta) \doteq \sup\{\|\boldsymbol{\phi}^x(u)-\boldsymbol{\phi}^x(u_1)\|\wedge\|\boldsymbol{\phi}^x(u_2)-\boldsymbol{\phi}^x(u)\| : u_1 \leq u \leq u_2, u_2-u_1 \leq \delta\}.$$

and $\|x\| \equiv |x_1| + |x_2|$. From the mean value theorem there exists a constant C such that for $u < v \in [a,b]$

$$\begin{aligned}\|\boldsymbol{\phi}^x(u) - \boldsymbol{\phi}^x(v)\| &\leq C(v-u)|T_x - \frac{1}{r}\log x|e^{\lambda_1 v|T_x - \frac{1}{r}\log x|}\\ &\leq C(v-u)e^{(\lambda_1+1)b|T_x-\frac{1}{r}\log x|},\end{aligned}$$

Thus if $\omega_x^{''}(\delta) \geq \varepsilon$ then

$$C\delta e^{(\lambda_1+1)b|T_x-\frac{1}{r}\log x|} \geq \varepsilon,$$

and therefore

$$\mathbb{P}(\omega_x^{''}(\delta) \geq \varepsilon) \leq \mathbb{P}\left(|T_x - \frac{1}{r}\log x| \geq b(\lambda+1)\log(\varepsilon/\delta)\right).$$

The condition (4.2) then follows by taking the limit as $x \to \infty$ (using (2.1)) and then sending δ to 0. □

4.1. *Crossover time.* We define the following stochastic time:

$$\xi \equiv \inf\{t \in (0,1] | Z_1(t) \geq Z_0(t)\},$$

which we refer to as the 'crossover' time, since it is the first time at which the Z_0 and Z_1 paths cross, and represents roughly the time at which the Z_1 or resistant cell population begins to dominate the tumor. In this section we investigate, using the limit theorems proven in the previous section, the distribution of the crossover time scaled by T_x. First we utilize the crossover time of the limit processes to obtain an estimate of this time. In particular, we define $\tilde{u}$ to be the solution to:

$$\phi_0(\tilde{u}, T_x - \frac{1}{r}\log x) = \phi_1(\tilde{u}, T_x - \frac{1}{r}\log x).$$

We obtain:

$$\tilde{u} = \frac{\log\left(\mu + (\lambda_1 + r)x^\alpha\right) - \log \mu}{T_x(\lambda_1 + r)}$$

and establish the following result.

THEOREM 5. *The estimate $\tilde{u}$ and the scaled crossover time, ξ/T_x, converge to each other in probability as $x \to \infty$, i.e., for all $\varepsilon > 0$*

$$\mathbb{P}(|\frac{\xi}{T_x} - \tilde{u}| > \varepsilon) \to 0 \text{ as } x \to \infty.$$

See section 7 for the proof of this result, which follows as an application of Theorem 3.

4.2. *Turnaround time: Progression of disease.* In this section we characterize the time at which the total tumor population stops declining and starts increasing. Define the following set of random times associated with the unscaled escape process:

$$\tau = \text{argmin}_{t \geq 0}\{Z_0(t) + Z_1(t)\}.$$

Using the sample path approximations we can approximate this set of times, rescaled by the extinction time T_x, as the random variable

$$u^* \equiv \frac{\log\left(\frac{r}{\lambda_1 \mu}\right) + \log\left(x^\alpha(\lambda_1 + r) - \mu\right)}{(\lambda_1 + r)T_x} \tag{4.3}$$

This corresponds to the time at which the approximated path of the total population size has derivative zero. Looking at the highest order terms in (4.3), we see that for large x,

$$u^* \approx \frac{\alpha r}{\lambda_1 + r}. \tag{4.4}$$

Thus a higher mutation rate, or smaller α, leads to a quicker turnaround time (relative to the extinction time). In addition, the decay rate r increases, the time of progression relative to the time of extinction increases.

Throughout this section we work with the sped-up but unscaled joint population processes, $Z_i(us_x(t))$. For simplicity write the sum of the mean of Z_0 and Z_1 as

$$\begin{aligned} f_{x,t}(u) &\equiv \mathbb{E}Z_0(us_x(t)) + \mathbb{E}Z_1(us_x(t)) \\ &= xe^{\lambda_0 u(\frac{1}{r}\log x + t)}\left(1 - \frac{\mu}{x^\alpha(\lambda_1 + r)}\right) + \frac{x^{1-\alpha}\mu e^{\lambda_1 u(\frac{1}{r}\log x + t)}}{\lambda_1 + r}. \end{aligned}$$

We will first show that with high probability, the critical point of $f_{x,t}$,

$$u^*(t) \equiv \frac{\log\left(\frac{r}{\lambda_1 \mu}\right) + \log\left(x^\alpha(\lambda_1 + r) - \mu\right)}{(\lambda_1 + r)\left(\frac{1}{r}\log x + t\right)}, \tag{4.5}$$

is the close to the minimum of $Z_0(us_x(t)) + Z_1(us_x(t))$. We then establish that this statement is in fact true uniformly for t in compact sets, and that τ/T_x is well approximated by u^*.

Since $u^*(t)$ is a critical point of $f_{x,t}$ we have the following representation that will be useful

$$f_{x,t}(u^*(t)) = xe^{\lambda_0 u^*(t)(\frac{1}{r}\log x + t)}\left(1 - \frac{\mu}{x^\alpha(\lambda_1 + r)}\right)\left(1 + \frac{r}{\lambda_1}\right),$$

and

$$\begin{aligned} &f_{x,t}(u^*(t) + y) \\ &\quad = xe^{\lambda_0 u^*(t)(\frac{1}{r}\log x + t)}\left(1 - \frac{\mu}{x^\alpha(\lambda_1 + r)}\right)\left(e^{\lambda_0 y(\frac{1}{r}\log x + t)} + \frac{r}{\lambda_1}e^{\lambda_1 y(\frac{1}{r}\log x + t)}\right). \end{aligned}$$

Therefore

$$f_{x,t}(u^*(t) + y) = f_{x,t}(u^*(t))\left[\left(\frac{\lambda_1}{\lambda_1 + r}\right)\left(x^{-y}e^{\lambda_0 yt} + \frac{r}{\lambda_1}x^{\lambda_1 y/r}e^{\lambda_1 yt}\right)\right]. \tag{4.6}$$

With this 'steepness' at the minimum property we can establish that with high probability (for x large) the minimum of the total population is achieved at $u^*(t)$.

Lemma 4. *For $\varepsilon > 0$,*

$$\mathbb{P}\left(\frac{\tau}{\frac{1}{r}\log x + t} \cap [u^*(t) - \varepsilon, u^*(t) + \varepsilon]^c \neq \emptyset\right) \to 0$$

as $x \to \infty$.

The proof of this result is deferred to section 7.

Similarly to the approximation result in Lemma 2, it is then possible to establish that an analogous result holds uniformly for t in compact sets.

Lemma 5. *For $\varepsilon > 0$ and a constant $M > 0$,*

(i)

$$\mathbb{P}(\sup_{t\in[-M,M]} \inf_{u\in[0,u^*(t)-\varepsilon)} Z_0(us_x(t)) + Z_1(us_x(t)) < Z_0(u^*T_x) + Z_1(u^*T_x)) \to 0.$$

(ii)

$$\mathbb{P}(\sup_{t\in[-M,M]} \inf_{u\in[0,u^*(t)+\varepsilon)} Z_0(us_x(t)) + Z_1(us_x(t)) < Z_0(u^*T_x) + Z_1(u^*T_x)) \to 0.$$

as $x \to \infty$.

See Section 7 for details of the proof.

We can now establish that the turnaround time of the scaled process τ normalized by the extinction time T_x converges in probability to u^*:

THEOREM 6. *For* $\varepsilon > 0$,

$$\mathbb{P}\left(\frac{\tau}{T_x} \cap [u^* - \varepsilon, u^* + \varepsilon]^c \neq \emptyset\right) \to 0$$

as $x \to \infty$.

PROOF. Using similar techniques as in the proof of Theorem 3 the result follows easily from the previous two lemmas. □

In Figure 4.2 we compare the sample probability density function of τ/T_x from simulations of the (Z_0, Z_1) process with the theoretical PDF of u^*. It is observed that even with an initial starting population of size $x = 100000$ the comparisons are favorable. Thus in the application of interest where x is on the order of 10^6 cells or greater we expect these limiting approximations to be of use.

5. An example: recurrence dynamics in Non-Small Cell Lung Cancer. In this section we apply the results to a simple model of drug resistance in non-small cell lung cancer (NSCLC). Non-small cell lung cancer is a disease in which malignant cells form in the tissues of the lung; it is the most common type of lung cancer, which causes over 150,000 deaths per year in the U.S. In recent years, a new class of targeted anti-cancer drugs called tyrosine kinase inhibitors has been developed. These inhibitors target molecules specifically within cancer cells and inhibit key signaling pathways such as the Epidermal Growth Factor Receptor (EGFR). Two such inhibitors, erlotinib and gefitinib, have been shown to be extremely successful in reducing tumor burden in a substantial subset of NSCLC patients.

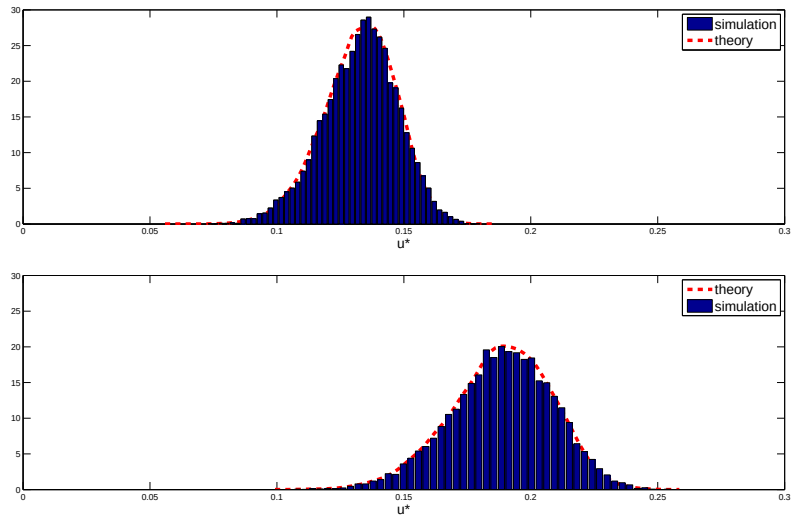

FIG 2. *Sample PDF of* τ/T_x *from simulation of* (Z_0, Z_1) *process compared with theoretical PDF of* u^**, for initial size* $x = 100000$ *and two parameter sets. Top:* $x = 100000, r_0 = 1.0, d_0 = 1.5, r_1 = 2.0, d_1 = 1.0, \mu = 0.01$*. Bottom:* $x = 100000, r_0 = 1.0, d_0 = 1.75, r_1 = 2.0, d_1 = 1.0, \mu = 0.01$*.*

However, point mutations in the binding site of the drug have been identified that confer resistance to both therapies, and thus lead to recurrence or progression of the disease.

In previous work [2] we characterized the *in vitro* growth rates of a pair of human NSCLC cell lines which were sensitive or resistant to the drug erlotinib (see Figure 5). Here we utilize this experimental growth kinetic data and apply our results on turnaround time distribution to study the properties of the time of disease progression. In particular, for a series of drug concentrations we characterize the distribution of the random time u^*, using the experimental data to ascertain r_0, d_0, r_1, and d_1. In addition, we use known estimates of the biological parameter $\mu_x \approx 10^{-8}$, which corresponds to the mutation probability per cell division per base pair in the genome [21, 16]. We can then apply our estimates of the turnaround time distribution to study how the time until progression varies as a function of drug concentration. These distributions of u^* are helpful in predicting the likely success of the therapy. In particular, u^* indicates the fraction of the total time that the drug is effective (T_x) at which disease progression occurs. If the distribution of u^* for a particular drug at a specific concentration has most of its mass bounded far below 1, the chance that the sensitive cell drug

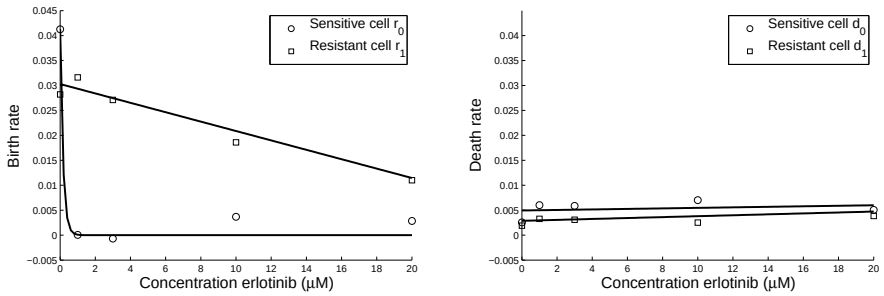

FIG 3. *Growth and death rate data (hours^{-1}) for erlotinib-sensitive (PC-9) and erlotinib-resistant NSCLC cells as a function of drug concentration (data published in [2]).*

population is eradicated by the time of progression is extremely low. On the other hand, drugs whose profiles which place most of the u^* distribution's mass closer to 1 have better prospects of eliminating the tumor. In Figure 5 we plot the u^* distribution for a NSCLC tumor starting with 10^9 sensitive cells treated with erlotinib at various concentrations. Note that the current standard of care, the FDA approved dose elicits a concentration of 3uM in the plasma which corresponds to the upper right plot. As the drug concentration increases, the distribution of u^* moves accordingly to the right; however, even at the highest concentration the majority of the mass is still bounded well below 1 which indicates likely failure of the therapy. In clinical observations, following an initial response in terms of tumor reduction, 100 percent of patients develop resistance usually within 24 months of starting treatment [18].

One major clinical question in NSCLC treatment today is: once the disease has progressed and the tumor size begins to increase, what course of therapy is optimal? In particular, should the drug be withdrawn or should the patient be kept on erlotinib or gefitinib? If drug is maintained, how long should it be administered beyond progression? Here, estimates of the u^* distribution can be of use. We note that τ is a clinically observable quantity since it represents the time until disease progression from the start of treatment. Once τ is observed, using Theorem 6 and the approximation in (4.4) we can approximate T_x, which represents the time at which the entire drug-sensitive population is eradicated. This gives a clear endpoint, T_x beyond which erlotinib therapy is unwarranted. Furthermore, we can easily obtain the distribution of the population size of resistant cells at this time $Z_1(T_x)$ to estimate the projected resistant tumor size at the time the sensitive cells are eradicated. This information aids in determining whether

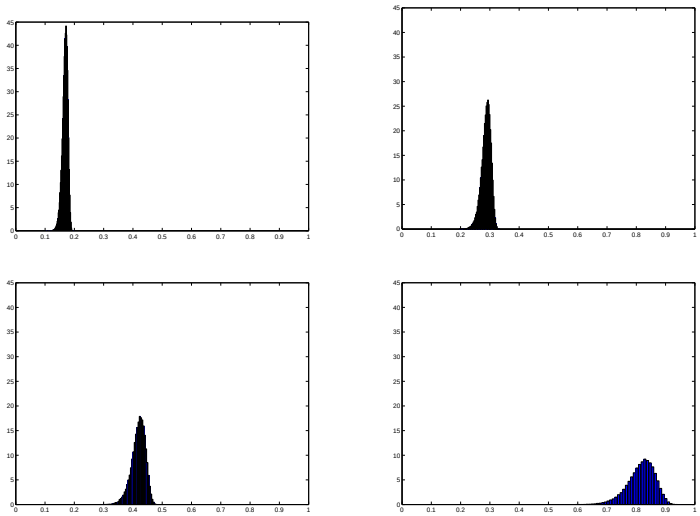

FIG 4. *Distributions of the turnaround time, u^*, for a NSCLC tumor with initially 10^9 sensitive cells, treated with erlotinib at $1\mu M$ (top left), $3\mu M$ (top right), $5\mu M$ (bottom left), and $10\mu M$ (bottom right).*

erlotinib treatment should be maintained until T_x or a switch to alternative therapy should be made prior to T_x.

6. Summary. In this work we have considered the stochastic dynamics of escape from extinction in a binary branching process model. By considering the large starting population limit, we approximate the birth death process with a simpler stochastic process whose only randomness is inherited from the weak limit of the extinction time. Using this limit, we approximate the distribution of the time until the total population begins to increase, and the time at which the escape mutants first begin to dominate the population. One of many possible future extensions is to consider the problems in this paper in a non-Markovian setting, i.e., non-exponential distribution between events. This work contributes to a growing body of literature concerned with the mathematical understanding of cancer evolution, as well as to the general understanding of extinction and escape paths in branching process process models. In future work we examine the setting $\alpha = 1$, where $O(1)$ mutations arise before extinction and escape from extinction is not assured in the large x limit.

7. Proof of Main Results.

7.1. *Proof of Theorem 2.* We first establish the scaling of the mean for large initial population x.

Proof. By conditioning on the path of Z_0 until T_x we get the formula (3.1), performing a change of measure and flipping the order of integration (by Tonelli's Thm) we see

$$\begin{aligned}\mathbb{E}[Z_1(vT_x)] &= \mu_x \mathbb{E}\left[\int_0^\infty 1_{\{T_x \geq y/v\}} Z_0(y) \exp\left(\lambda_1(vT_x - y)\right) dy\right] \\ &= \mu_x \int_0^\infty \mathbb{E}\left[1_{\{T_x \geq y/v\}} Z_0(y) \exp\left(\lambda_1(vT_x - y)\right)\right] dy \\ &= \mu_x \int_0^\infty \int_{y/v}^\infty \sum_{n=1}^\infty n e^{\lambda_1(vt-y)} \mathbb{P}\left(T_x \in dt | Z_0(y) = n\right) \mathbb{P}(Z_0(y) = n) dy.\end{aligned}$$

Note that $\mathbb{P}(T_x \in dt | Z_0(y) = n) = g_n(t-y)dt$, where g_n is the density of the extinction time for a population starting from a population of size n, and can be written as $g_n(t) = n\left(G(t)\right)^{n-1} g(t)$, where g is the density of the extinction time for a population starting from a single cell, and G is the cdf. Therefore upon rearranging the order of integration we get that

$$\begin{aligned}&\mathbb{E}[Z_1(vT_x)] \\ &\quad = \mu_x \int_0^\infty \int_{y/v}^\infty e^{\lambda_1(vt-y)} g(t-y) \left(\sum_{n=1}^\infty n^2 \left(G(t-y)\right)^{n-1} \mathbb{P}(Z_0(y) = n)\right) dt dy.\end{aligned}$$

Next define

$$F_x(s,t) = \mathbb{E}[s^{Z_0(t)}] \text{ and } F(s,t) = \mathbb{E}[s^{Z_0(t)} | Z_0(0) = 1, Z_1(0) = 0],$$

and observe that due to the independence of the branching structure $F_x(s,t) = (F(s,t))^x$. Therefore,

$$\begin{aligned}&\sum_{n=1}^\infty n^2 s^{n-1} \mathbb{P}(Z_0(t) = n) \\ &\quad = s \frac{\partial^2}{\partial s^2} F_x(s,t) + \frac{\partial}{\partial s} F_x(s,t) \\ &\quad = sx\left(F(s,t)\right)^{x-1} \frac{\partial^2}{\partial s^2} F(s,t) + sx(x-1)\left(F(s,t)\right)^{x-2} \left(\frac{\partial}{\partial s} F(s,t)\right)^2 \\ &\quad\quad + x\left(F(s,t)\right)^{x-1} \frac{\partial}{\partial s} F(s,t). \qquad (7.1)\end{aligned}$$

For ease of notation we will simply write $\frac{\partial}{\partial s}F(s,t) = F'(s,t)$. Using (7.1) we obtain:

$$\mathbb{E}[Z_1(vT_x)] =$$
$$x\mu_x \int_0^\infty \int_{y/v}^\infty e^{\lambda_1(vt-y)} g(t-y)\left(F(G(t-y),y)\right)^{x-2}\Big[(x-1)G(t-y)\left(F'(G(t-y),y)\right)^2$$
$$+ G(t-y)F(G(t-y),y)F''(G(t-y),y) + F(G(t-y),y)F'(G(t-y),y)\Big]\,dydt.$$

This expression can be analyzed using techniques from [10]. In particular if we introduce the change of variable

$$t = \frac{1}{r}\left(z + \log cx\right) = z_x,$$

and observe that

$$e^{\lambda_1(vz_x - y)} = e^{\lambda_1(vz/r-y)}(cx)^{\lambda_1 v/r}.$$

After the change of variables we have

$$\mathbb{E}[Z_1(vT_x)] = I_1(x,v) + I_2(x,v)$$

where

$$I_1(x,v) = \frac{x(x-1)(cx)^{\frac{\lambda_1 v}{r}}\mu_x}{r}\int_0^\infty \int_{\frac{ry}{v}-\log cx} \Phi_1(z,y)dzdy$$
$$I_2(x,v) = \frac{x(cx)^{\frac{\lambda_1 v}{r}}\mu_x}{r}\int_0^\infty \int_{\frac{ry}{v}-\log cx}^\infty \Phi_2(z,y)dzdy$$

and

$$\Phi_1(z,y)$$
$$= e^{\lambda_1(vz/r-y)} g(z_x-y)G(z_x-y)\left(F(G(z_x-y),y)\right)^{x-2}\left(F'(G(z_x-y),y)\right)^2$$
$$\Phi_2(z,y)$$
$$= e^{\lambda_1(vz/r-y)} g(z_x-y)\left(F(G(z_x-y),y)\right)^{x-1}\big[G(z_x-y)F''(G(z_x-y),y)$$
$$+ F'(G(z_x-y),y)\big]$$

We will now establish that for $v \in (0,1]$, $x^{\alpha-1-\frac{\lambda_1 v}{r}} I_1(x,v) = \tilde{I}_1(x,v) \to I_1(v)$ and $x^{\alpha-1-\frac{\lambda_1 v}{r}} I_2(x,v) = \tilde{I}_2(x,v) \to 0$ as $x \to \infty$. The integrand of $\tilde{I}_1(x,v)$ is

$$f_x(z,y)$$
$$= c_0 x g(z_x-y)G(z_x-y)\left(F(G(z_x-y),y)\right)^{x-2}\left(F'(G(z_x-y),y)\right)^2 e^{\lambda_1(vz/r-y)},$$

where $c_0 = \mu c^{\lambda_1 v/r}/r$. From [10] we know that as $z \to \infty$, $g(z) \sim rce^{-rz}$, and therefore

$$g(z_x - y) \sim \frac{r}{x} e^{ry} e^{-z}, \tag{7.2}$$

as $x \to \infty$. Next note that there exists a $\xi_x \in (G(z_x - y), 1)$ such that

$$\begin{aligned} F'(G(z_x - y), y) &= F'(1, y) + (1 - G(z_x - y))\, F''(\xi_x, y) \\ &= e^{-ry} + O\left(1 - G(z_x - y)\right), \end{aligned}$$

and therefore

$$F'(G(z_x - y), y) \sim e^{-ry}. \tag{7.3}$$

Lastly observe that

$$F(G(z_x - y), y) = 1 + e^{-ry}\left(G(z_x - y) - 1\right) + O\left(G(z_x - y) - 1\right)^2,$$

and therefore

$$\log F(G(z_x - y), y) = -\left(1 - G(z_x - y)\right) e^{-ry} + O\left(G(z_x - y) - 1\right)^2.$$

Observe that

$$1 - G(z_x - y) \sim \frac{e^{ry} e^{-z}}{x},$$

which gives that

$$\left(F(G(z_x - y), y)\right)^{x-2} \sim \exp[-e^{-z}]. \tag{7.4}$$

Combining (7.2),(7.3), and (7.4) we see that

$$\lim_{x \to \infty} f_x(z, y) = c_1 e^{-z(1 - \frac{\lambda_1}{r} v)} e^{-y(r + \lambda_1)} \exp[-e^{-z}], \tag{7.5}$$

where $c_1 = rc_0$. In order to evaluate the limit of $\tilde{I}_1$ it thus remains to show that the limit can be passed inside the integral, this will be done by finding an integrable function h such that $f_x(z, y) \le h(z, y)$. First note that since $G(z) \le 1$, and $F'(s, t) \le E_1 Z_0(t) = e^{-rt}$ we have

$$f_x(z, y) \le c_0 x e^{\lambda_1 (vz/r - y)} g(z_x - y)\left(F(G(z_x - y), y)\right)^{x-2} e^{-2ry}.$$

Then observe that there exists a constant k_1 such that

$$g(z_x - y) \le k_1 e^{-r(z_x - y)} = \frac{k_2}{x} e^{-z} e^{ry}. \tag{7.6}$$

Since $\log x \leq x-1$ we have

$$\begin{aligned}(F(G(z_x-y),y))^{x-2} &= \exp\left[(x-2)\log F(G(z_x-y),y)\right] \\ &\leq \exp\left[-(x-2)(1-F(G(z_x-y),y))\right].\end{aligned}$$

Using results from the proofs of Proposition 1 and Lemma 1 of [10], we can establish that

$$1-F(G(z_x-y),y) \geq \frac{e^{-z}}{x}. \tag{7.7}$$

Based on (7.6) and (7.7) we see that we can use the dominating function

$$h(z,y) = k_3 e^{\lambda_1(\frac{vz}{r}-y)} e^{-z} e^{-ry} \exp[-k_2 e^{-z}].$$

With this result we see that

$$\begin{aligned}\lim_{x\to\infty} \tilde{I}_1(x,v) &= c_0 \int_0^\infty \int_{-\infty}^\infty \lim_{x\to\infty} f_x(z,y)dzdy \\ &= c_1 \int_0^\infty \int_{-\infty}^\infty e^{\lambda_1(\frac{vz}{r}-y)} e^{-yr} e^{-z} \exp[-e^{-z}]dzdy \\ &= \frac{\mu c^{\lambda_1 v/r}\Gamma(1-\frac{\lambda_1 v}{r})}{(\lambda_1+r)}.\end{aligned}$$

We now consider $\tilde{I}_2$, first observe that for $(s,t) \in [0,1]\times[0,\infty)$ there exists finite k_4 such that $F'(s,t) \leq k_4$ and $F''(s,t) \leq k_4$, and of course $F(s,t) \leq 1$. Therefore if we consider $\tilde{I}_2$ in terms of the original variables, there exists a finite constant k_5 such that

$$\begin{aligned}\tilde{I}_2(x,v) &\leq k_5 x^{-v\lambda_1/r} \int_0^\infty \int_{y/v}^\infty e^{\lambda_1(vt-y)} g(t-y)dtdy \\ &= k_5 x^{-v\lambda_1/r} \int_0^\infty \int_{y(1/v-1)}^\infty e^{-\lambda_1 y} g(s) e^{\lambda_1 v(s+y)} dsdy = k_6 x^{-v\lambda_1/r},\end{aligned}$$

where the first equality follows by using the change of variable $s = t-y$. Thus $\tilde{I}_2(x,v) \to 0$ as $x \to \infty$ for $v \in (0,1]$. □

7.2. *Proof of Lemma 1.* We will start by establishing item (ii). Define $\mathcal{F}_\infty^0$ to be the sigma algebra generated by the wave 0 population until their eventual extinction, then

$$E\left[Z_1(s_x(t))|\mathcal{F}_\infty^0\right] = \mu_x \int_0^{s_x(t)} Z_0(y) e^{\lambda_1(s_x(t)-y)} dy$$

and therefore

$$\mathbb{E}\left[Z_0(s)Z_1(s_x(t))\right] = e^{\lambda_1 s_x(t)}\mu_x \int_0^{s_x(t)} \mathbb{E}[Z_0(y)Z_0(s)]e^{-\lambda_1 y}dy$$

Now we establish the second moment result, item (i). For simplicity we evaluate $\mathbb{E}[Z_1(t)]$ for a positive t. For ease of notation we will use the following $\tilde{\mathbb{E}}[\cdot] = \mathbb{E}[\cdot|\mathcal{F}^0_\infty]$. Consider a partition of $[0,t]$, $0 < \Delta < 2\Delta < \ldots < t$, where $\Delta = t/m$ for a large integer m. Then we can write

$$\begin{aligned}
\mathbb{E}[Z_1(t)^2|\mathcal{F}^0_\infty] &= \tilde{\mathbb{E}}\left(\sum_{j=0}^{m}\sum_{k=1}^{N_j} B_{j,k}(t-\tau_{j,k})\right)^2 \\
&= \sum_{j=0}^{m}\tilde{\mathbb{E}}\left(\sum_{k=1}^{N_j} B_{j,k}(t-\tau_{j,k})\right)^2 \\
&\quad + \sum_{j=0}^{m}\sum_{\ell\neq j}\tilde{\mathbb{E}}\left(\sum_{k=1}^{N_j} B_{j,k}(t-\tau_{j,k})\right)\tilde{\mathbb{E}}\left(\sum_{k=1}^{N_\ell} B_{\ell,k}(t-\tau_{\ell,k})\right),
\end{aligned}$$

where N_j is the number of type-1 mutants created in $[j\Delta,(j+1)\Delta)$, $\{B_{j,k}\}$ is a collection of i.i.d binary birth death processes with birth rate a_1 and death rate d_1, and $\tau_{j,k}$ is the time of creation for the kth mutant created in $[j\Delta,(j+1)\Delta)$. In the previous display we have used the independence of the branching process to derive the second equality. For $0 \le j \le m$,

$$\begin{aligned}
\mathbb{E}N_j &= \Delta\mu_x Z_0(j\Delta) + o(\Delta) \\
\mathbb{E}N_j^2 &= \Delta\mu_x Z_0(j\Delta)(1+\Delta\mu_x Z_0(j\Delta)) + o(\Delta).
\end{aligned}$$

Therefore

$$\begin{aligned}
\tilde{\mathbb{E}}\left(\sum_{k=1}^{N_j} B_{j,k}(t-\tau_{j,k})\right)^2 &= \tilde{\mathbb{E}}N_j\tilde{\mathbb{E}}B(t-\tau_j)^2 + \tilde{\mathbb{E}}[N_j(N_j-1)](\tilde{\mathbb{E}}B(t-\tau_j))^2 \\
&= \Delta\mu_x Z_0(j\Delta)\tilde{\mathbb{E}}B(t-\tau_j)^2 + O(\Delta^2) \\
&= \Delta\mu_x Z_0(j\Delta)\left(\frac{2r_1}{\lambda_1}e^{2\lambda_1(t-\tau_j)} - \frac{r_1+d_1}{\lambda_1}e^{\lambda_1(t-\tau_j)}\right) \\
&\quad + O(\Delta^2),
\end{aligned}$$

and

$$\begin{aligned}
\tilde{\mathbb{E}}\left(\sum_{k=1}^{N_\ell} B_{\ell,k}(t-\tau_{\ell,k})\right) &= \Delta\mu_x Z_0(j\Delta)\tilde{\mathbb{E}}B(t-\tau_j) \\
&= \Delta\mu_x Z_0(j\Delta)e^{\lambda_1(t-\tau_j)}.
\end{aligned}$$

Using the previous two expressions we get

$$\tilde{\mathbb{E}}[Z_1(t)^2] = \Delta\mu_x \sum_{j=0}^{m} Z_0(j\Delta)\left(\frac{2r_1}{\lambda_1}e^{2\lambda_1(t-\tau_j)} - \frac{r_1+d_1}{\lambda_1}e^{\lambda_1(t-\tau_j)}\right)$$
$$+ (\mu_x\Delta)^2 \sum_{j=0}^{m}\sum_{\ell=0,\ell\neq j}^{m} Z_0(j\Delta)Z_0(\ell\Delta)e^{\lambda_1(t-\tau_j)}e^{\lambda_1(t-\tau_\ell)}.$$

Sending $\Delta \to 0$, integrating over Z_0, and replacing t with $\frac{1}{r}\log x + t$ gives us the desired formula for item (i).

Item (iii) follows immediately from item (i).

7.3. *Proof of Lemma 2.*

Proof. We will prove the more difficult second statement first. We observe that it suffices to prove that as $x \to \infty$,

$$\mathbb{P}\left(\sup_{u\in[0,1]} e^{-\lambda_1 ut}\left|Z_1^x(us_x(t)) - \phi_1^x(u,t)\right| > e^{-\lambda_1 t}\varepsilon\right) \to 0.$$

Next observe that

$$\frac{e^{-\lambda_1 ut}}{x^{-\alpha+1+\lambda_1 u/r}}\left(Z_1(us_x(t)) - \phi_1(u,t)\right)$$
$$= \left(\frac{e^{-\lambda_1 ut}}{x^{1+\lambda_1 u/r}}Z_1(us_x(t)) - \frac{\mu}{x^{1+\alpha}}\int_0^{us_x(t)} Z_0(s)e^{-\lambda_1 s}ds\right)x^\alpha$$
$$+ \frac{\mu}{x}\int_0^{us_x(t)}(Z_0(s) - xe^{\lambda_0 s})e^{-\lambda_1 s}ds,$$

and therefore

$$\mathbb{P}\left(\sup_{u\in[0,1]} e^{-\lambda_1 ut}\left|Z_1^x(us_x(t)) - \phi_1^x(u,t)\right| > e^{-\lambda_1 t}\varepsilon\right)$$
$$\leq \mathbb{P}\left(\sup_{u\in[0,1]} x^\alpha\left|\frac{e^{-\lambda_1 ut}}{x^{1+\lambda_1 u/r}}Z_1(us_x(t)) - \frac{\mu}{x^{1+\alpha}}\int_0^{us_x(t)} Z_0(s)e^{-\lambda_1 s}ds\right| > \varepsilon/2\right)$$
$$+ \mathbb{P}\left(\sup_{u\in[0,1]} \frac{\mu}{x}\int_0^{us_x(t)}\left|Z_0(s) - xe^{\lambda_0 s}\right|e^{-\lambda_1 s}ds > \varepsilon/2\right).$$

However we can observe that the process considered in second expression in the sum is monotonic in u, and the process considered in the first expression

is a martingale in u, which allows us to arrive at the following simpler inequality:

$$\begin{aligned}
&\mathbb{P}\left(\sup_{u\in[0,1]} e^{-\lambda_1 ut}\left|Z_1^x(us_x(t)) - \phi_1^x(u,t)\right| > e^{-\lambda_1 t}\varepsilon\right) \\
&\leq \frac{4x^{2\alpha}}{\varepsilon^2}\mathbb{E}\left[\left(\frac{e^{-\lambda_1 t}}{x^{\lambda_1/r+1}}Z_1(s_x(t)) - \frac{\mu}{x^{1+\alpha}}\int_0^{s_x(t)} Z_0(s)e^{-\lambda_1 s}ds\right)^2\right] \\
&\quad + \mathbb{P}\left(\frac{\mu}{x}\int_0^{s_x(t)} |Z_0(s) - xe^{\lambda_0 s}|e^{-\lambda_1 s}ds > \varepsilon/2\right). \qquad (7.8)
\end{aligned}$$

Consider the latter quantity first, where it suffices to show that as $x \to \infty$,

$$\frac{\mu}{x}\int_0^{s_x(t)} \mathbb{E}\left[\left|Z_0(s) - xe^{\lambda_0 s}\right|\right] e^{-\lambda_1 s}ds \to 0.$$

Next observe that

$$\mathrm{Var}(Z_0(s)) = x\left(\frac{r_0+d_0}{\lambda_0}\right)\left(e^{2\lambda_0 s} - e^{\lambda_0 s}\right).$$

It follows from the Cauchy-Schwarz inequality that

$$\frac{\mu}{x}\int_0^{s_x(t)} \mathbb{E}\left[\left|Z_0(s) - xe^{\lambda_0 s}\right|\right] e^{-\lambda_1 s}ds = O(x^{-1/2}).$$

Moving onto the first term in (7.8),

(7.9)

$$\begin{aligned}
&x^{2\alpha}\mathbb{E}\left[\left(\frac{e^{-\lambda_1 t}}{x^{\lambda_1/r+1}}Z_1(s_x(t)) - \frac{\mu}{x^{1+\alpha}}\int_0^{s_x(t)} Z_0(s)e^{-\lambda_1 s}ds\right)^2\right] \\
&= \left(\frac{e^{-\lambda_1 t}x^{\alpha}}{x^{1+\lambda_1/r}}\right)^2 \mathbb{E}\left[Z_1(s_x(t))^2\right] - 2\frac{x^{2\alpha}\mu e^{-\lambda_1 t}}{x^{2+\lambda_1/r+\alpha}}\int_0^{s_x(t)} e^{-\lambda_1 s}\mathbb{E}\left[Z_0(s)Z_1(s_x(t))\right]ds \\
&\quad + \left(\frac{\mu}{x}\right)^2\int_0^{s_x(t)}\int_0^{s_x(t)} \mathbb{E}[Z_0(s)Z_0(y)]e^{-\lambda_1 s}e^{-\lambda_1 y}dsdy.
\end{aligned}$$

Using Lemma 1 we see that (7.9) can be written as

$$\frac{2r_1x^{\alpha-1}\mu}{\lambda_1}\int_0^{s_x(t)} e^{s(\lambda_0-2\lambda_1)}ds - \frac{(r_1+d_1)\mu e^{-\lambda_1 t}}{\lambda_1 x^{1-\alpha+\lambda_1/r}}\int_0^{s_x(t)} e^{s(\lambda_0-\lambda_1)}ds = O\left(x^{\alpha-1}\right),$$

thus establishing the result (ii).

We now move onto the proof of item (i). First observe that

$$\sup_{u\in[0,a]} |Z_0^x(us_x(t)) - \phi_0^x(u,t)| \le e^{-\lambda_0 a t^-} \sup_{u\in[0,a]} e^{-\lambda_0 u t} |Z_0^x(us_x(t)) - \phi_0^x(u,t)|,$$

where $t^- = -\min(t,0)$. We will show that as $x \to \infty$,

$$\mathbb{P}\left(\sup_{u\in[0,a]} \left|e^{-\lambda_0 u t} Z_0^x(us_x(t)) - 1\right| \ge \varepsilon e^{\lambda_0 a t^-}\right) \to 0.$$

Observe that, $e^{-\lambda_0 u t} x^{u-1} Z_0\left(us_x(t)\right)) - 1$ is a martingale with respect to u. Therefore it suffices to show that as $x \to \infty$,

$$\mathbb{E}\left[e^{-\lambda_0 a t} Z_0^x(as_x(t)) - 1\right]^2 = x^{2a-1} e^{-2a\lambda_0 t} \mathrm{Var}_1 Z_0(as_x(t)) \to 0,$$

where $\mathrm{Var}_1 Z_0(t)$ represents the variance of $Z_0(t)$ starting with an initial population size 1. The previous expression reduces to,

$$x^{a-1} e^{-a\lambda_0 t}(x^{-a} e^{-\lambda_0 a t} - 1)(r_0 + d_0)/\lambda_0,$$

and since we have assumed that $a < 1$, the result is established. □

7.4. *Proof of Lemma 3.*

Proof. Throughout the proof assume that $x > e^{rM}$. We first establish the result for the Z_1 population by showing the following monotonicity property in t:

$$\begin{aligned} &\sup_{u\in[0,1]} \left|Z_1^x\left(u\left(\frac{1}{r}\log x + t_0\right)\right) - \phi_1^x\left(u, t_0\right)\right| \\ &\qquad \le \sup_{u\in[0,1]} \left|Z_1^x\left(u\left(\frac{1}{r}\log x + t_1\right)\right) - \phi_1^x\left(u, t_1\right)\right| \end{aligned} \tag{7.10}$$

for $t_0 \le t_1$. For any $u \in [0,1]$ set

$$\bar{u} \equiv \frac{u\left(\frac{1}{r}\log x + t_0\right)}{\frac{1}{r}\log x + t_1},$$

which of course implies that $\bar{u}\left(\frac{1}{r}\log x + t_1\right) = u\left(\frac{1}{r}\log x + t_0\right)$. In addition, observe that $\bar{u} \le u$ and thus $x^{-\lambda_1 u/r} \le x^{-\lambda_1 \bar{u}/r}$. Therefore,

$$\begin{aligned} &x^{-\lambda_1 u/r}\left|Z_1\left(u\left(\frac{1}{r}\log x + t_0\right)\right) - \phi_1\left(u, t_0\right)\right| \\ &\qquad \le x^{-\lambda_1 \bar{u}/r}\left|Z_1\left(\bar{u}\left(\frac{1}{r}\log x + t_1\right)\right) - \phi_1\left(\bar{u}, t_1\right)\right|. \end{aligned}$$

Since $u \in [0,1]$ was arbitrary we have that

$$\sup_{t\in[-M,M]} \sup_{u\in[0,1]} x^{\alpha-1-\lambda_1 u/r} \left|Z_1\left(us_x(t)\right) - \phi_1\left(u,t\right)\right|$$
$$\leq \sup_{u\in[0,1]} x^{\alpha-1-\lambda_1 u/r} \left|Z_1\left(u\left(\frac{1}{r}\log x + M\right)\right) - \phi_1\left(u,M\right)\right|.$$

Result (ii) now follows by an application of Lemma 2.

The proof of result (i) will follow a similar approach. In particular for $t \leq M$, and $u \in [0,a]$ define

$$\hat{u} = u\left(\frac{\frac{1}{r}\log x + t}{\frac{1}{r}\log x + M}\right).$$

Notice that

$$u - \hat{u} = u\left(\frac{M-t}{\frac{1}{r}\log x + M}\right) \leq a\left(\frac{M-t}{\frac{1}{r}\log x + M}\right) = n(x,M).$$

Using the definition of $n(x,M)$ it follows that $n(x,M)\log x \leq 2arM$ which implies that $x^{u-\hat{u}} \leq e^{2arM}$. Based on the definition of $\hat{u}$ and the upper bound on $x^{u-\hat{u}}$

$$x^u \left|Z_0(us_x(t)) - \phi_0(u,t)\right|$$
$$= x^{u-\hat{u}}x^{\hat{u}}\left|Z_0^x\left(\hat{u}\left(\frac{1}{r}\log x + M\right)\right) - \phi_0(\tilde{u},M)\right|$$
$$\leq e^{2arM}x^{\hat{u}}\left|Z_0\left(\hat{u}\left(\frac{1}{r}\log x + M\right)\right) - \phi_0(\hat{u},M)\right|.$$

Since the previous inequality holds for any u, we know that for any $t \in [-M,M]$,

$$\sup_{u\in[0,a]} x^u \left|Z_0(us_x(t)) - \phi_0(u,t)\right|$$
$$\leq e^{2arM} \sup_{u\in[0,a]} x^u \left|Z_0\left(u\left(\frac{1}{r}\log x + M\right)\right) - \phi_0(u,M)\right|.$$

Thus the result of (i) is established by using the result of Lemma 2 for $t = M$. □

7.5. *Proof of Theorem 5.*

PROOF. First we prove that $\mathbb{P}(\xi/T_x \leq \tilde{u}-\epsilon) \to 0$ as $x \to \infty$. In particular recall that

$$\phi_0(u,t) = x^{1-u}e^{\lambda_0 ut}$$
$$\phi_1(u,t) = \frac{\mu x^{1-\alpha+\lambda_1 u/r}e^{\lambda_1 ut}}{\lambda_1 + r}\left(1 - e^{u(\lambda_0-\lambda_1)t}x^{(\lambda_0-\lambda_1)u/r}\right).$$

Then, let us utilize the notation $d(T_x) \equiv T_x - \frac{1}{r}\log x$ to represent the deviation of T_x from its scaling:

$$\begin{aligned}
&\mathbb{P}\left(\sup_{u\leq\tilde{u}-\epsilon}(Z_1(uT_x) - Z_0(uT_x)) > 0\right)\\
&\quad\leq \mathbb{P}\Bigg(\sup_{u\leq\tilde{u}-\epsilon} x^{u-1}(Z_1(uT_x) - \phi_1(u,d(T_x)) + (\phi_1(u,d(T_x)) - \phi_0(u,d(T_x)))\\
&\qquad + (\phi_0(u,d(T_x)) - Z_0^x(uT_x)) > 0, |d(T_x)| \leq \frac{1}{r}\log x\Bigg)\\
&\qquad + \mathbb{P}\left(|d(T_x)| > \frac{1}{r}\log x\right)\\
&\quad\leq \mathbb{P}\Bigg(\sup_{u\leq\tilde{u}-\epsilon} x^{u-1}(Z_1(uT_x) - \phi_1(u,d(T_x)))\\
&\qquad + \sup_{u\leq\tilde{u}-\epsilon} x^{u-1}(\phi_1(u,d(T_x)) - \phi_0(u,d(T_x)))\\
&\qquad + \sup_{u\leq\tilde{u}-\epsilon} x^{u-1}(\phi_0(u,d(T_x)) - Z_0^x(uT_x)) > 0, |d(T_x)| \leq \frac{1}{r}\log x\Bigg)\\
&\qquad + \mathbb{P}\left(|d(T_x)| > \frac{1}{r}\log x\right).
\end{aligned}$$

Clearly $\mathbb{P}\left(|d(T_x)| > \frac{1}{r}\log x\right) \to 0$ as $x \to \infty$, and it thus remains to analyze the first expression on the right hand side of previous display. First notice that if $|d(T_x)| \leq \frac{1}{r}\log x$ then

$$\begin{aligned}
&\sup_{u\leq\tilde{u}-\epsilon} x^{u-1}(\phi_1(u,d(T_x)) - \phi_0(u,d(T_x)))\\
&\qquad = x^{\tilde{u}-\epsilon-1}(\phi_1(\tilde{u}-\epsilon,d(T_x)) - \phi_0(\tilde{u}-\epsilon,d(T_x))).
\end{aligned}$$

Therefore it suffices to show that the following converges to 0,

$$\mathbb{P}\Bigg(\sup_{u\le\tilde{u}-\epsilon} x^{u-1}(Z_1(uT_x)-\phi_1(u,d(T_x))) + x^{\tilde{u}-\epsilon-1}(\phi_1(\tilde{u}-\epsilon,d(T_x))-\phi_0(\tilde{u}-\epsilon,d(T_x)))+ + \sup_{u\le\tilde{u}-\epsilon} x^{u-1}(\phi_0(u,d(T_x))-Z_0^x(uT_x))>0\Bigg). \tag{7.11}$$

To study this let us start by considering the first term in the sum above:

$$\begin{aligned}\sup_{u\le\tilde{u}-\epsilon}&|x^{u-1}(Z_1(uT_x)-\phi_1(u,d(T_x)))| \\ &\le \sup_{u\le\tilde{u}-\epsilon} x^{u(1+\frac{\lambda_1}{r})-\alpha}\sup_{u\le\tilde{u}-\epsilon}|x^{\alpha-\lambda_1 u/r-1}(Z_1(uT_x)-\phi_1(u,d(T_x))|.\end{aligned} \tag{7.12}$$

The second term in the product converges to zero in probability via Theorem 3. The first term tends to zero by the following argument:

$$\begin{aligned}\log\left[\sup_{u\le\tilde{u}-\epsilon} x^{u(1+\frac{\lambda_1}{r})-\alpha}\right] &= \log\left[x^{(\tilde{u}-\epsilon)(1+\frac{\lambda_1}{r})-\alpha}\right] \\ &= \log\left[x^{-(1+\frac{\lambda_1}{r})\epsilon-\alpha}\exp[\tilde{u}(1+\frac{\lambda_1}{r})\log x]\right] \\ &= \left(-\alpha-(1+\frac{\lambda_1}{r})\epsilon\right)\log x+\frac{\log x}{rT_x}\log\left[1+x^\alpha\left(\frac{\lambda_1+r}{\mu}\right)\right] \\ &\le \left(-\alpha-(1+\frac{\lambda_1}{r})\epsilon\right)\log x+\frac{\log x}{rT_x}\log\left[2x^\alpha\left(\frac{\lambda_1+r}{\mu}\right)\right] \\ &= \alpha\frac{\log x}{rT_x}(\log x-rT_x)-\left(\frac{\lambda_1}{r}+1\right)\epsilon\log x+\frac{\log x}{rT_x}\log\left[\frac{2(\lambda_1+1)}{\mu}\right],\end{aligned}$$

where in the third equality we have utilized the fact that

$$\tilde{u}\left(\frac{\lambda_1+r}{r}\right)=\frac{1}{rT_x}\log\left[\frac{(\lambda_1+r)x^\alpha}{\mu}+1\right]$$

due to the definition of $\tilde{u}$. Observe that $-\left(\frac{\lambda_1}{r}+1\right)\epsilon\log x$ diverges to negative infinity, while the first and third terms approach finite limits. This can be seen by observing that

$$\frac{\log x}{rT_x}\to 1$$

in probability. Thus, we conclude that (7.12) goes to zero in probability.

The second term in (7.11) is considered next. Via the definition of $\phi_0(u, d(T_x))$ and the limit result on the extinction time (2.1), we have that as $x \to \infty$

$$x^{\tilde{u}-\epsilon-1}\phi_0(\tilde{u}-\epsilon, d(T_x)) \Rightarrow c^{-(\tilde{u}-\varepsilon)}e^{-(\tilde{u}-\varepsilon)\eta},$$

where η is a standard Gumbel random variable and c is the positive Yaglom constant. Importantly this limit random variable is positive with probability one. The first term can be shown to approach zero by noting that

$$x^{\tilde{u}-\epsilon-1}\phi_1(\tilde{u}-\epsilon, d(T_x)) = x^{(\tilde{u}-\epsilon)(1+\frac{\lambda_1}{r})-\alpha}x^{\alpha-(\tilde{u}-\epsilon)\lambda_1/r-1}\phi_1(\tilde{u}-\epsilon, d(T_x))$$

where the first term approaches zero, as argued previously since its log aproaches negative infinity, and the product of the remaining terms approaches a constant times the exponential of a Gumbel, which is again a result of (2.1). The third term in (7.11)

$$\sup_{u\le\tilde{u}-\epsilon} x^{u-1}(\phi_0(u, d(T_x)) - Z_0^x(uT_x))$$

converges to zero in probability by Theorem 3. Therefore

$$\limsup_{x\to\infty} \mathbb{P}(\xi/T_x \le \tilde{u}-\epsilon) \le \mathbb{P}(c^{-(\tilde{u}-\varepsilon)}e^{-(\tilde{u}-\varepsilon)\eta} \le 0) = 0.$$

Next, we need to show that $\mathbb{P}(\xi/T_x \ge \tilde{u}+\epsilon) \to 0$. We have by definition of ξ, that

$$\begin{aligned}
\mathbb{P}(\xi/T_x > \tilde{u}+\epsilon) &\le \mathbb{P}(Z_0((\tilde{u}+\epsilon)T_x) - Z_1((\tilde{u}+\epsilon)T_x) > 0) \\
&= \mathbb{P}(x^{\alpha-(\tilde{u}+\epsilon)\frac{\lambda_1}{r}-1}\left(Z_0((\tilde{u}+\epsilon)T_x) - Z_1((\tilde{u}+\epsilon)T_x)\right) > 0) \\
&= \mathbb{P}\Bigg(x^{\alpha-(\tilde{u}+\epsilon)\frac{\lambda_1}{r}-1}(Z_0((\tilde{u}+\epsilon)T_x) - \phi_0(\tilde{u}+\epsilon, d(T_x))) \\
&\quad + x^{\alpha-(\tilde{u}+\epsilon)\frac{\lambda_1}{r}-1}(\phi_0(\tilde{u}+\epsilon, d(T_x)) - \phi_1(\tilde{u}+\epsilon, d(T_x))) \\
&\quad + x^{\alpha-(\tilde{u}+\epsilon)\frac{\lambda_1}{r}-1}(\phi_1(\tilde{u}+\epsilon, d(T_x)) - Z_1((\tilde{u}+\epsilon)T_x)) > 0\Bigg)
\end{aligned}$$

It is easily shown that the RHS of the previous display goes to 0 using analogous arguments from the analysis of (7.11). □

7.6. *Proof of Lemma 4.* Here we establish Lemma 4, namely that $u^*(t)$ approximates $\tau/(\frac{1}{r}\log x + t)$.

Proof. We will prove first that

$$\mathbb{P}\left(\inf_{u\in[u^*(t)+\varepsilon,\infty]} Z_1(us_x(t)) < Z_0(u^*(t)s_x(t)) + Z_1(u^*(t)s_x(t))\right) \to 0. \tag{7.13}$$

Consider the following decomposition of the event of interest,

$$\begin{aligned}
&\mathbb{P}\left(\inf_{u\in[u^*(t)+\varepsilon,\infty]} Z_1(us_x(t)) < Z_0(u^*(t)s_x(t)) + Z_1(u^*(t)s_x(t))\right)\\
&\leq \mathbb{P}\left(\inf_{u\in[u^*(t)+\varepsilon,\infty]} Z_1(us_x(t)) < f(u^*(t)+\varepsilon/2)\right)\\
&\quad + \mathbb{P}\left(Z_0(u^*(t)s_x(t)) + Z_1(u^*(t)s_x(t)) > f_{x,t}(u^*(t)+\varepsilon/2)\right).
\end{aligned}$$

We can apply Markov's inequality to the last probability to see

$$\begin{aligned}
&\mathbb{P}\left(Z_0(u^*(t)s_x(t)) + Z_1(u^*(t)s_x(t)) > f_{x,t}(u^*(t)+\varepsilon/2)\right)\\
&\quad\leq \frac{\mathbb{E}[Z_0(u^*(t)s_x(t)) + Z_1(u^*(t)s_x(t))]}{f_{x,t}(u^*(t)+\varepsilon/2)}\\
&\quad= \frac{f_{x,t}(u^*(t),t)}{f_{x,t}(u^*(t)+\varepsilon/2)}\\
&\quad= O(x^{-\lambda_1\varepsilon/2r}),
\end{aligned}$$

where the last equality follows from the 'steepness' at the minimum property (4.6).

We can do a further decomposition,

$$\begin{aligned}
&\mathbb{P}\left(\inf_{u\in[u^*(t)+\varepsilon,\infty]} Z_1(us_x(t)) < f(u^*(t)+\varepsilon/2)\right)\\
&= \mathbb{P}\left(\inf_{u\in[u^*(t)+\varepsilon,\infty]} Z_1(us_x(t)) < f(u^*(t)+\varepsilon/2), Z_1\left((u^*(t)+\varepsilon)s_x(t)\right) < f_{x,t}(u^*(t)+3\varepsilon/4)\right)\\
&\quad + \mathbb{P}\left(\inf_{u\in[u^*(t)+\varepsilon,\infty]} Z_1(us_x(t)) < f(u^*(t)+\varepsilon/2), Z_1\left((u^*(t)+\varepsilon)s_x(t)\right) > f_{x,t}(u^*(t)+3\varepsilon/4)\right)\\
&\leq \mathbb{P}\left(Z_1\left((u^*(t)+\varepsilon)s_x(t)\right) < f_{x,t}(u^*(t)+3\varepsilon/4)\right)\\
&\quad + \mathbb{P}\left(\inf_{u\in[u^*(t)+\varepsilon,\infty]} Z_1(us_x(t)) < f(u^*(t)+\varepsilon/2), Z_1\left((u^*(t)+\varepsilon)s_x(t)\right) > f_{x,t}(u^*(t)+3\varepsilon/4)\right)
\end{aligned} \tag{7.14}$$

Using Chebyshev's inequality and the result in (4.6) again, we see that

$$\begin{aligned}
&\mathbb{P}\left(Z_1\left((u^*(t)+\varepsilon)s_x(t)\right) < f_{x,t}(u^*(t)+3\varepsilon/4)\right)\\
&\quad= \mathbb{P}(Z_1\left((u^*(t)+\varepsilon)s_x(t)\right) - \mathbb{E}[Z_1\left((u^*(t)+\varepsilon)s_x(t)\right)]\\
&\qquad\qquad < f_{x,t}(u^*(t)+3\varepsilon/4) - \mathbb{E}[Z_1\left((u^*(t)+\varepsilon)s_x(t)\right)])\\
&\quad\leq \frac{\mathbb{E}[|Z_1\left((u^*(t)+\varepsilon)s_x(t)\right) - \mathbb{E}Z_1\left((u^*(t)+\varepsilon)s_x(t)\right)|^2]}{|f_{x,t}(u^*(t)+\frac{3\epsilon}{4}) - \mathbb{E}Z_1\left((u^*(t)+\varepsilon)s_x(t)\right)|^2}.
\end{aligned} \tag{7.15}$$

Let us consider first the denominator in the above expression, and note that since $u^*(t)$ minimizes $\mathbb{E}Z_0(us_x(t)) + \mathbb{E}Z_1(us_x(t))$, we have that

$$\frac{r}{\lambda_1} x e^{\lambda_0 u^*(t)(\frac{1}{r}\log x + t)} \left(1 - \frac{\mu}{x^\alpha(\lambda_1 + r)}\right) = \frac{x^{1-\alpha}\mu}{\lambda_1 + r} e^{\lambda_1 u^*(t)(\frac{1}{r}\log x + t)}.$$

Thus,

$$\begin{aligned}
&\mathbb{E}[Z_1\left((u^*(t)+\varepsilon)s_x(t)\right)] \\
&\quad = x e^{\lambda_0 u^*(t)(\frac{1}{r}\log x + t)} \left[-\frac{\mu}{x^\alpha(\lambda_1 + r)} e^{\lambda_0 \varepsilon(\frac{1}{r}\log x + t)} \right. \\
&\qquad \left. + \frac{r}{\lambda_1}\left(1 - \frac{\mu}{x^\alpha(\lambda_1 + r)}\right) e^{\lambda_1 \varepsilon(\frac{1}{r}\log x + t)} \right].
\end{aligned}$$

Also,

$$\begin{aligned}
&f_{x,t}(u^*(t) + 3\varepsilon/4) = \\
&x e^{\lambda_0 u^*(t)(\frac{1}{r}\log x + t)} \left(1 - \frac{\mu}{x^\alpha(\lambda_1 + r)}\right) (e^{\lambda_0 3\varepsilon/4(\frac{1}{r}\log x + t)} + (r/\lambda_1) e^{\lambda_1 3\varepsilon/4(\frac{1}{r}\log x + t)}),
\end{aligned}$$

and therefore

$$\begin{aligned}
&|f_{x,t}(u^*(t) + \frac{3\epsilon}{4}) - \mathbb{E}Z_1\left((u^*(t)+\varepsilon)s_x(t)\right)|^2 \\
&= \left| x^{1-u^*(t)} e^{\lambda_0 u^*(t) t} \left[\left(1 - \frac{\mu}{x^\alpha(\lambda_1 + r)}\right) \left(x^{-3\varepsilon/4} e^{3\lambda_0 \varepsilon t/4} + \frac{r}{\lambda_1} x^{3\lambda_1 \varepsilon/4 r} e^{3\lambda_1 \varepsilon t/4} \right) \right. \right. \\
&\quad \left. \left. + \frac{\mu}{x^\alpha(\lambda_1 + r)} x^{-\varepsilon} e^{-\varepsilon t} - \frac{r}{\lambda_1}\left(1 - \frac{\mu}{x^\alpha(\lambda_1 + r)}\right) x^{\lambda_1 \varepsilon/r} e^{\lambda_1 \varepsilon t} \right] \right|^2 \\
&= \Omega\left(x^{2(\frac{\lambda_1 \varepsilon}{r} + 1 - u^*(t))} \right).
\end{aligned}$$

Next we consider the variance term. For ease of notation define $\theta(t) \equiv (u^*(t) + \varepsilon)(\frac{1}{r}\log x + t)$, then from item (iii) of Lemma 1

$$\begin{aligned}
&\mathrm{Var}[Z_1\left((u^*(t)+\varepsilon)s_x(t)\right)] \\
&\quad = \left(\frac{\mu}{x^\alpha}\right)^2 \int_0^{\theta(t)} \int_0^{\theta(t)} \mathrm{Cov}(Z_0(s), Z_0(y)) e^{\lambda_1(2\theta(t) - s - y)} ds dy \\
&\qquad + \left(\frac{\mu}{x^\alpha}\right) \int_0^{\theta(t)} \mathbb{E}Z_0(s) \left(2r_1 e^{2\lambda_1(\theta(t) - s)} - (r_1 + d_1) e^{\lambda_1(\theta(t) - s)}\right) ds
\end{aligned}$$

We now establish that $\mathrm{Cov}(Z_0(s), Z_0(y))$ is an $O(x)$ quantity. Since the population Z_0 starts with x independent cells we can write the covariance

as

$$\begin{aligned}\mathrm{Cov}(Z_0(s), Z_0(y)) &= \mathbb{E}\left[\sum_{j=1}^{x} Z_0^{(j)}(y) \sum_{i=1}^{x} Z_0^{(i)}(s)\right] - \mathbb{E}[Z_0(s)]\mathbb{E}[Z_0(y)] \\ &= x\mathbb{E}_1[Z_0(s)Z_0(y)] + x(x-1)\mathbb{E}_1[Z_0(s)]\mathbb{E}_1[Z_0(y)] - \mathbb{E}[Z_0(s)]\mathbb{E}[Z_0(y)] \\ &= x\mathrm{Cov}_1(Z_0(s), Z_0(y)) = O(x).\end{aligned}$$

Based on this we know that the second term in the definition of $\mathrm{Var}Z_1$ is the dominant term and therefore $\mathrm{Var}[Z_1((u^*(t)+\varepsilon)s_x(t))] = O\left(x^{1-\alpha+2\lambda_1\frac{u^*(t)+\varepsilon}{r}}\right)$

Then, in order to establish that (7.15) goes to zero it suffices to show that

$$\left(1-\alpha+2\lambda_1\frac{u^*(t)+\varepsilon}{r} - 2(\frac{\lambda_1\varepsilon}{r} + 1 - u^*(t))\right)\log x \to -\infty.$$

The result in the previous display follows from the definition of $u^*(t)$ in (4.5), and therefore we can conclude that

$$\lim_{x\to\infty} \mathbb{P}\left(Z_1\left((u^*(t)+\varepsilon)s_x(t)\right) < f_{x,t}(u^*(t)+3\varepsilon/4)\right) = 0.$$

It remains to show that the final probability in display (7.14) is negligible for large x. Observe that if we start out with a collection of n independent cells each following branching processes with net-growth rate $\lambda_1 > 0$, then by the law of large numbers the fraction, f_n, of those cells whose lineage eventually dies out satisfies the following limit: $\lim_{n\to\infty} f_n = p_E(\lambda_1) < 1$, where $p_E(\lambda_1)$ is the probability of a single cell's descendants going extinct and is strictly less than 1 because $\lambda_1 > 0$. Therefore define $\rho_x(u,t)$ to be the fraction of type-1 cells present at time $u(\frac{1}{r}\log x + t)$ that eventually die out. Notice then that in order for the event described in the last line of display (7.14) to occur it is necessary that

$$\rho_x(u^*(t)+\varepsilon, t) \geq 1 - \frac{f_{x,t}(u^*(t)+\varepsilon/2)}{f_{x,t}(u^*(t)+3\varepsilon/4)}.$$

Then from the 'steepness' property we have that for x sufficiently large $\rho_x(u^*(t)+\varepsilon,t) > p_E(\lambda_1)+\eta$, for some $\eta > 0$. Of course from the law of large numbers we have that

$$\lim_{x\to\infty} \mathbb{P}\left(\rho_x(u^*(t)+\varepsilon,t) > p_E(\lambda_1)+\eta, Z_1\left((u^*(t)+\varepsilon)s_x(t)\right) > f_{x,t}(u^*(t)+3\varepsilon/4)\right) = 0,$$

thus establishing (7.13).

Moving on we next establish that:

$$\mathbb{P}\left(\inf_{u\in[0,u^*(t)-\varepsilon]} Z_0(us_x(t)) < Z_0(u^*(t)s_x(t)) + Z_1(u^*(t)s_x(t))\right) \to 0 \tag{7.16}$$

as $x \to \infty$. Note that based on arguments from the above case it suffices to establish that the following probability converges to 0 as $x \to \infty$:

$$\begin{aligned}
&\mathbb{P}\left(\inf_{u\in[0,u^*(t)-\varepsilon]} Z_0(us_x(t)) < f_{x,t}(u^*(t)-\varepsilon/2)\right)\\
&= \mathbb{P}\left(\inf_{u\in[0,u^*(t)-\varepsilon]} Z_0(us_x(t)) - c_0 f_{x,t}(u^*(t)-\varepsilon/2) < (1-c_0) f_{x,t}(u^*(t)-\varepsilon/2)\right)\\
&= \mathbb{P}\left(\sup_{u\in[0,u^*(t)-\varepsilon]} (c_0 f_{x,t}(u^*(t)-\varepsilon/2) - Z_0(us_x(t))) > (c_0-1) f_{x,t}(u^*(t)-\varepsilon/2)\right)\\
&\le \frac{1}{(c_0-1)f_{x,t}(u^*(t)-\varepsilon/2)}\mathbb{E}\left[(c_0 f_{x,t}(u^*(t)-\varepsilon/2) - Z_0\left((u^*(t)-\varepsilon)s_x(t)\right))^+\right]\\
&\le \frac{c_0 f_{x,t}(u^*(t)-\varepsilon/2)}{(c_0-1)f_{x,t}(u^*(t)-\varepsilon/2)}\mathbb{P}\left(Z_0\left((u^*(t)-\varepsilon)s_x(t)\right) < c_0 f_{x,t}(u^*(t)-\varepsilon/2)\right)
\end{aligned}$$

where we chose $c_0 > 1$, and the penultimate inequality follows from Doob's inequality and that $-Z_0(\cdot s_x(t))$ is a submartingale. The final probability can be rewritten as

$$\begin{aligned}
\mathbb{P}\big(&Z_0\left((u^*(t)-\varepsilon)s_x(t)\right) - \mathbb{E}Z_0\left((u^*(t)-\varepsilon)s_x(t)\right)\\
&< c_0 f_{x,t}(u^*(t)-\varepsilon/2) - \mathbb{E}Z_0\left((u^*(t)-\varepsilon)s_x(t)\right)\big).
\end{aligned}$$

Note since $\mathbb{E}Z_0\left((u^*(t)-\varepsilon)s_x(t)\right) = x^{1+\varepsilon-u^*(t)}e^{\lambda_0 t(u^*(t)-\varepsilon)}$ and $f_{x,t}(u^*(t)-\varepsilon/2) = x^{1+\varepsilon/2-u^*(t)}(1+o(1))$, we know that there exists a positive constant C_0 such that for x sufficient large

$$\mathbb{E}Z_0\left((u^*(t)-\varepsilon)s_x(t)\right) - cf_{x,t}(u^*(t)-\varepsilon/2) \ge C_0 x^{1+\varepsilon-u^*(t)},$$

and therefore for x sufficiently large

$$\begin{aligned}
&\mathbb{P}\big(Z_0\left((u^*(t)-\varepsilon)s_x(t)\right) - \mathbb{E}Z_0\left((u^*(t)-\varepsilon)s_x(t)\right)\\
&\qquad < cf_{x,t}(u^*(t)-\varepsilon/2) - \mathbb{E}Z_0\left((u^*(t)-\varepsilon)s_x(t)\right)\big)\\
&\le \mathbb{P}\left(\left|Z_0\left((u^*(t)-\varepsilon)s_x(t)\right) - \mathbb{E}Z_0\left((u^*(t)-\varepsilon)s_x(t)\right)\right| > C_0 x^{1+\varepsilon-u^*(t)}\right).
\end{aligned}$$

Thus, via Chebyshev's inequality we have

$$\begin{aligned}
&\mathbb{P}\left(|Z_0\left((u^*(t)-\varepsilon)s_x(t)\right)-\mathbb{E}Z_0\left((u^*(t)-\varepsilon)s_x(t)\right)|>C_0x^{1+\varepsilon-u^*(t)}\right)\\
&\leq \frac{x^{1/2}\left(\mathrm{Var}_1 Z_0\left((u^*(t)-\varepsilon)s_x(t)\right)\right)^{1/2}}{C_0x^{1+\varepsilon-u^*(t)}}\\
&= O\left(x^{-(1+\varepsilon-u*)/2}\right),
\end{aligned}$$

where the final equality follows by evaluating the variance of $Z_0(u^*(t)(\frac{1}{r}\log x+t))$. □

7.7. *Proof of Lemma 5.*

Proof. As in the proof of the previous lemma, we consider the deviations to the left and right of $u^*(t)$ separately. First, note that if $t_0 > t_1$, then

$$\begin{aligned}
&\inf_{u\in[0,u^*(t_0)-\varepsilon)} Z_0(u(\frac{1}{r}\log x+t_0))+Z_1(u(\frac{1}{r}\log x+t_0))\\
&=\inf\{Z_0(s)+Z_1(s): s\leq (u^*(t_0)-\varepsilon)(\frac{1}{r}\log x+t_0)\}\\
&\geq \inf\{Z_0(s)+Z_1(s): s\leq (u^*(t_1)-\varepsilon)(\frac{1}{r}\log x+t_1)\}\\
&=\inf_{u\in[0,u^*(t_1)-\varepsilon)} Z_0(u(\frac{1}{r}\log x+t_1))+Z_1(u(\frac{1}{r}\log x+t_1)).
\end{aligned}$$

Furthermore, it follows from the definition of $u^*(t)$ that $Z_i(u^*(t)(\frac{1}{r}\log x+t)) = Z_i(u^*(s)(\frac{1}{r}\log x+s))$ for all s,t. In particular $Z_i\left(u^*(s)(\frac{1}{r}\log x+s)\right) = Z_i(u^*T_x)$. Then via Lemma (4)

$$\mathbb{P}\left(\sup_{t\in[-M,M]}\inf_{u\in[0,u^*(t)-\varepsilon)} Z_0(us_x(t))+Z_1(us_x(t))<Z_0(u^*T_x)+Z_1(u^*T_x)\right)\to 0.$$

as $x\to\infty$. A similar argument can be used for deviations to the right of $u^*(t)$ to show that

$$\mathbb{P}\left(\sup_{t\in[-M,M]}\inf_{u\in[u^*(t)+\varepsilon,\infty)} Z_0(us_x(t))+Z_1(us_x(t))<Z_0(u^*T_x)+Z_1(u^*T_x)\right)\to 0$$

as $x\to\infty$, establishing the lemma. □

References.

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

[2] 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.

[3] A. Coldman and J. Goldie. A stochastic model for the origin and treatment of tumors containing drug-resistant cells. *Bull. Math. Biol.*, 48(3/4):279–292, 1986.

[4] J. Foo and F. Michor. Evolution of resistance to targeted anti-cancer therapies during general dosing schedules. *Journal of Theoretical Biology*, 263:179–188, 2010.

[5] J.H. Goldie and A.J. Coldman. A mathematical model for relating the drug sensitivity of tumours to their spontaneous mutation rate. *Cancer Treat Rep*, 63:1727–1733, 1979.

[6] J.H. Goldie and A.J. Coldman. Quantitative model for multiple levels of drug resistance in clinical tumors. *Cancer Treat Rep*, 67:923–931, 1983.

[7] L. Harnevo and Z. Agur. Drug resistance as a dynamic process in a model for multistep gene amplification under various levels of selection stringency. *Cancer Chemotherapy Pharmacology*, 30:469–476, 1992.

[8] L. Harnevo and Z. Agur. The dynamics of gene amplification described as a multi type compartmental model and as a branching process. *Mathematical Biosciences*, 103:115–138, 1992.

[9] Y. Iwasa, F. Michor, and M.A. Nowak. Evolutionary dynamics of escape from biomedical intervention. *Proc. Roy. Soc. Lond. B*, 270:2572–2578, 2003.

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

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

[12] N.L. Komarova. Stochastic modeling of drug resistance in cancer. *Journal of Theoretical Biology*, 239:351–366, 2006.

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

[14] J. Michaelson, E. Halpern, and D. Kopans. Breast cancer: Computer simulation method for estimating optimal intervals for screening. *Radiology*, 212:551–560, 1999.

[15] F. Michor, M.A. Nowak, and Y. Iwasa. Evolution of resistance to cancer therapy. *Curr Pharm Design*, 12:261–271, 2006.

[16] A. R. Oller, P. Rastogi, S. Morgenthaler, and W. G. Thilly. A statistical model to estimate variance in long term-low dose mutation assays: Testing of the model in a human lymphoblastoid mutation assay. *Mutat. Res.*, 216:149–161, 1989.

[17] A.G. Pakes. Asymptotic results for the extinction time of Markov branching processes allowing for emigration, I. Random Walk Decrements. *Advances in Applied Probability*, 21:243–269, 1989.

[18] W. Pao and J. Chmielecki. Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer. *Nature Reviews Cancer*, 10:760–774, 2010.

[19] S. Sagitov and M.C. Serra. Multitype Bienayme-Galton-Watson processes escaping extinction. *Advances in Applied Probability*, 41:225–246, 2009.

[20] M. Sehl, H. Zhou, J. Sinsheimer, and K. Lange. Extinction models for cancer stem cell therapy. *Mathematical Biosciences*, 234:132–146, 2011.

[21] R. Seshadri, R. J. Kutlaca, K. Trainor, C. Matthews, and A. A. Morley. Mutation rate of normal and malignant human lymphocytes. *Cancer Res*, 263:407–409, 1987.

[22] X. Wu and M. Kimmel. A note on the path to extinction of critical Markov branching processes. *Statistics and Probability Letters*, 80:263–269, 2010.

University of Minnesota
School of Mathematics
206 Church St SE
Minneapolis, MN 55455
E-mail: jyfoo@math.umn.edu

University of Minnesota
Industrial and Systems Engineering
111 Church St SE
Minneapolis, MN 55455
E-mail: kevin.leder@isye.umn.edu