Dynamics of cancer recurrence

Jasmine Foo ∗1 and Kevin Leder †2

1School of Mathematics, University of Minnesota
2Industrial and Systems Engineering, University of Minnesota

January 3, 2012

Abstract

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, during each replication of a sensitive cell, a mutation arise with small probability, conferring drug-resistance (and a net positive growth rate) to the mutant daughter cell. If such 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

∗jyfoo@math.umn.edu
†kevin.leder@isye.umn.edu
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 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 \[7, 8\]. Nowak and collaborators 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 the ‘turnaround time’ which we define as the time that 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 the sample distribution of the turnaround time. 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 easily extended to study more general offspring distributions, and thus may be useful for studying escape dynamics in viral populations, for instance.

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 $x \to \infty$. 
Figure 1: Sample simulation of escape dynamics (population size vs time). Black line: total population size of the tumor, Red line: resistant cell population size, Blue line: sensitive cell population size. The magenta dot 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 sample distribution of the turnaround time 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 $\mu = 0.01$. 

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 with probability \( \mu \) during each division of a sensitive cell, and each of these mutations give rise to a supercritical Markovian binary branching process initialized by one mutant cell with net growth rate \( \lambda_1 > 0 \). The total population population of mutants, which we will call ‘resistant cells,’ is denoted \( Z_1(t) \). These processes are defined on a probability space \((\Omega, \mathcal{F}, P)\). In addition denote the filtration defined by \( Z_i(\cdot) \) up until time \( t \) by \( \mathcal{F}_t^i \). 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

\[
T_x - \frac{1}{r} \log x \Rightarrow \frac{1}{r} (\eta + \log c), \text{ as } x \to \infty
\]

where \( \eta \) 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 \( u \in [0, 1) \),

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

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 \( v > 0 \),

\[
E[Z_1(vT_x) \mid Z_0(s), s \leq T_x] = E \left[ r_0 \mu_1 T_x \int_0^{v\lambda_1} Z_0(uT_x) \exp (\lambda_1 T_x(v - u)) \, du \right].
\]
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 heuristically estimate the expected value as

$$E[Z(vT_x)] = \frac{r_0 \mu}{r} \log x \int_0^{v \land 1} x^{1-u} x^{\lambda_1 \lambda_0 (v-u)} du$$

$$= \frac{r_0 \mu}{r} x^{1-\lambda_0 v} \log x \int_0^{v \land 1} x^{-u(1+\frac{\lambda_1}{\lambda_0})} du$$

$$= \frac{r_0 \mu}{\lambda_1 - \lambda_0} x^{1+\lambda_1 v} \left( 1 - \exp \left[ -(v \land 1)(1 + \frac{\lambda_1}{\lambda_0}) \log x \right] \right).$$

Thus we observe that this expected value is finite for all $v > 0$. However, if we retain the assumption that $Z_0$ decays exponentially but now approximate the extinction time with the following distribution (based on the weak limit (1))

$$T_x = -\log(xc) + \eta/\lambda_0,$$

and therefore $Z_0(uT_x) = x^{1-u} e^{-u \eta}$, where as usual $\eta$ is a standard Gumbel. We then have

$$E[Z(vT_x)] = \frac{r_0 \mu x}{r} \int_{-\infty}^{\infty} \exp \left[ -e^{-z} - z \right] \log(xc) + z \int_0^{v \land 1} \left( xc \right)^{-u} e^{-uz} e^{\frac{\lambda_1}{\lambda_0} (v-u)(\log(xc) + z)} du dz$$

$$= \frac{r_0 \mu x}{r} \int_{-\infty}^{\infty} \exp \left[ -e^{-z} - z \right] \log(xc) + z \exp \left[ -\frac{\lambda_1}{\lambda_0} v \log(xc) + z \right]$$

$$\int_0^{v \land 1} \exp \left[ -u(1 - \frac{\lambda_1}{\lambda_0})(\log(xc) + z) \right] du dz$$

$$= \frac{(xc)^{-v \lambda_1 / r}}{r} \int_{-\infty}^{\infty} e^{-z} \exp[-e^{-z}] e^{\frac{\lambda_1}{\lambda_0} vz} \left( 1 - e^{-(v \land 1)(1 - \frac{\lambda_1}{\lambda_0})(\log(xc) + z)} \right) dz.$$

If we assume for convenience that $v \leq 1$ then applying the change of variable $y = e^{-z}$ gives

$$= \frac{x r_0 \mu(xc)^{v \lambda_1 / r}}{r} \int_0^{\infty} y^{-\lambda_1 v / r} e^{-y} dy - \frac{x r_0 \mu \exp(xc)^{-v}}{r} \int_0^{\infty} y^{v} e^{-y} dy.$$

This expression has the same polynomial rate of growth in $x$ as in the approximation of using the deterministic $T_x$ estimate; however, here the expression diverges to infinity for $v > \frac{\lambda_1}{\lambda_0}$. Thus, we conclude that using a deterministic estimate for $T_x$ and neglecting the tails of the extinction time distribution can lead to a significant loss of information.

The previous heuristic argument is easily made rigorous, and in fact we can establish a stronger result. In particular,

**Proposition 1** Let $v \in [0,1)$ and $\frac{\lambda_1 v}{r} > 1$ then for all $x$ $E[Z_1(vT_x)]$ is infinite.
Thus if \( \lambda_1 u > -\lambda_0 \), the mean of the resistant cell population is infinite. In other words, at the extinction time of the sensitive cells \( (u = 1) \), if \( \lambda_1 \geq -\lambda_0 \) the resistant cell population has infinite mean. If the opposite is true \( (-\lambda_0 > \lambda_1) \) (i.e. the decay of the sensitive cells is faster than the growth rate of the resistant cells) then the resistant cell population has finite mean at the time of sensitive cell extinction. Of course this result holds even when \( x > 1 \) since the infinite mean is a result of the combination of the tails of the extinction time and the growth properties of the resistant cells.

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

\[
E[Z_1(vT_x)] = E \left[ r_0 \mu T_x \int_0^v Z_0(uT_x) \exp [\lambda_1 T_x(v-u)] \, du \right]
\]

\[
= r_0 \mu \int_0^v \int_0^\infty t e^{\lambda_1 t(v-u)} E[Z_0(uT_x) \mid T_x \in dt] g_x(t) \, dt \, du,
\]

where \( g_x(t) \, dt = \mathbb{P}(T_x \in dt) \). Noting the fact that if \( u < 1 \) then \( E[Z_0(uT_x) \mid T_x \in dt] > 1 \), we can bound this from below by

\[
E[Z_1(vT_x)] \geq r_0 \mu \int_0^v \int_0^\infty t e^{\lambda_1 t(v-u)} g_x(t) \, dt \, du
\]

\[
= r_0 \mu \int_0^v \left( E[T_x] + \lambda_1 (v-u) \int_0^\infty e^{\lambda_1 s(v-u)} \int_s^\infty t g_x(t) \, dt \, ds \right) \, du
\]

\[
\geq r_0 \mu \int_0^v \left( E[T_x] + \lambda_1 (v-u) \int_0^\infty e^{\lambda_1 s(v-u)} \int_s^\infty g_x(t) \, dt \, ds \right) \, du
\]

\[
\geq r_0 \mu \int_0^v \left( E[T_x] + c \lambda_1 (v-u) \int_0^\infty e^{\lambda_1 s(v-u)} se^{-rs} \, ds \right) \, du.
\]

The penultimate inequality is based on the observation that the tail probability of the extinction time starting from one cell is smaller than the tail probability of the extinction time starting from \( x \) cells. In particular, define \( \tilde{F}_x(s) = \mathbb{P}(T_x > s) \) to be the tail probability at \( s \) of the extinction time of the sensitive cell process starting from size \( x \), and observe

\[
\tilde{F}_x(s) = 1 - \mathbb{P}(T_x \leq s) = 1 - (\mathbb{P}(T_1 \leq s))^x.
\]

Since \( (\mathbb{P}(T_x \leq s))^x \leq \mathbb{P}(T_1 \leq s) \), we have that \( \tilde{F}_x(s) \geq 1 - \mathbb{P}(T_1 \leq s) = \tilde{F}_1(s) \). The final inequality is based on 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 \( \lambda_1 v - r - \lambda_1 u > 0 \) and hence the integral

\[
\int_{t_0}^\infty se^{s(\lambda_1 v - r - \lambda_1 u)} \, ds
\]

diverges to \( \infty \). □
We can easily find the asymptotic growth rate of $E[Z_1(vT_x)]$ 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 $v \leq -\lambda_0/\lambda_1$. Earlier heuristic calculations indicate that the mean of $Z_1(vT_x)$ grows like $x^{1+\nu\lambda_1/r}$ as $x \to \infty$. In particular we have the following,

**Theorem 2** Assume that $r \geq \lambda_1$ then for $v \in (0,1]$ we have that

$$E[Z_1(vT_x)] \sim x^{1+\nu\lambda_1/r} \frac{r_0\mu (1 - \frac{\lambda_1 v}{r})}{b^{\lambda_1 v/r} (\lambda_1 + r)}.$$ 

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. Here, the mutation rate $\mu$ may decay to zero with increasing starting population, i.e. $\mu_x = \mu x^{-\alpha}$ where $\alpha \in [0,1)$. The setting where $\alpha \geq 1$ will be considered in later work.

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

$$Z_0^x(us_x(t)) = x^{u-1} Z_0\left( u \left( \frac{1}{r} \log x + t \right) \right)$$

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

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.

$$\phi_0^x(u,t) = E[Z_0^x(us_x(t))] = e^{\lambda_0 u t}$$

$$\phi_1^x(u,t) = E[Z_1^x(us_x(t))] = \frac{r_0\mu e^{\lambda_1 u t}}{\lambda_1 - \lambda_0} \left( 1 - e^{(\lambda_0 - \lambda_1)u t} \frac{\lambda_0 - \lambda_1}{r} \right).$$

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

$$\phi_i(u,t) = E[Z_i(us_x(t))].$$

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. 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 $\phi^x_t$ evaluated at the random parameter $T_x - \frac{1}{r} \log x$. In particular, we begin with the following lemma establishing the approximation at a fixed time point.

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

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

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

The proof, which can be found in section 7, proceeds via martingale arguments and the analysis of the second moments of the sensitive and resistant cells.

We can in fact strengthen the previous lemma by showing the convergence above is in fact uniform for $t$ in a compact set.

**Lemma 2** For $a \in (0, 1)$, $\varepsilon > 0$, and $M > 0$,

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

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

The proof follows by establishing that the probabilities in the statement of the lemma are monotone in the parameter $t$. Again we defer the proof until Section 7.

With this uniform approximation result, we are now able to establish our approximation for the paths evaluated at multiples of their extinction time. In particular,

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

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

(ii) \[ \lim_{x \to \infty} \mathbb{P} \left( \sup_{u \in [0, 1]} x^{1-u/1-1} \left| Z_1(uT_x) - \phi_1 \left( u, T_x - \frac{1}{r} \log x \right) \right| > \varepsilon \right) = 0. \]
Proof. We will only prove the result for $Z_0$, since the proof for $Z_1$ is identical. First observe that for finite positive $M$,
\[
\mathbb{P} \left( \sup_{u \in [0,a]} x^{u-1} \left| Z_0(uT_x) - \phi_0 \left( u, T_x - \frac{1}{r} \log x \right) \right| > \varepsilon \right)
\leq \mathbb{P} \left( T_x - \frac{1}{r} \log x \in [-M, M], \sup_{u \in [0,a]} x^{u-1} \left| Z_0(uT_x) - \phi_0 \left( u, T_x - \frac{1}{r} \log x \right) \right| > \varepsilon \right)
+ \mathbb{P} \left( \left| T_x - \frac{1}{r} \log x \right| > M \right).
\]
From (1) we know that for any $\delta > 0$ there exists $M_\delta < \infty$ such that
\[
\limsup_{x \to \infty} \mathbb{P} \left( \left| T_x - \frac{1}{r} \log x \right| > M_\delta \right) \leq \delta.
\]
Then via Lemma 2,
\[
\limsup_{x \to \infty} \mathbb{P} \left( T_x - \frac{1}{r} \log x \in [-M_\delta, M_\delta], \sup_{u \in [0,a]} x^{u-1} \left| Z_0(uT_x) - \phi_0 \left( u, T_x - \frac{1}{r} \log x \right) \right| > \varepsilon \right)
\leq \limsup_{x \to \infty} \mathbb{P} \left( \sup_{t \in [-M_\delta,M_\delta]} \sup_{u \in [0,a]} x^{u-1} \left| Z_0(usx(t)) - \phi_0(u, t) \right| > \varepsilon \right) = 0.
\]
Since $\delta > 0$ is arbitrary, $(i)$ is established.  

For what follows we define the following functions, if $u \in (0,1]$ then
\[
\psi_0(u) = e^{-u(\eta + \log c)}
\]
\[
\psi_1(u) = \frac{\tau_0 \mu}{\lambda + r} e^{\frac{\lambda_1 u}{r}(\eta + \log c)},
\]
and in addition define $\psi_0(0) = 1$ and $\psi_1(0) = 0$. These functions represent the limiting functions of our scaled population processes, however note that $\psi_1$ 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 $[0, b]$. This is of course a result of the massive influx of mutations near $t = 0$ in the unscaled process. If we were in the setting where $\alpha \geq 1$ it is our belief that the processes limit would be right continuous at the origin, however that is the topic of further work.

**Corollary 4** For $\alpha \in [0, 1)$ and $0 < a < b < 1$ the joint process
\[
(x^{-\frac{1}{r}}Z_0(T_x), x^{\alpha - \frac{\lambda_1}{r} - 1}Z_1(T_x)) \Rightarrow (\psi_0(\cdot), \psi_1(\cdot)),
\]
as $x \to \infty$ in the standard Skorokhod topology, $D([a,b])$. 

10
Proof. For ease of notation throughout this proof we will use the following notation
\[
\phi^x(u) = \left( \phi_0^x(u, T_x - \frac{1}{r} \log x), \phi_1^x(u, T_x - \frac{1}{r} \log x) \right).
\]

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 observe via the continuous mapping theorem and the weak convergence result in (1) we have finite dimensional convergence. It thus remains to establish tightness. Since our limit functions are continuous at \( u = a \) and \( u = b \), it suffices to establish that for every \( \varepsilon > 0 \)
\[
\lim_{\delta \to 0} \lim_{x \to \infty} \sup \mathbb{P}(\omega''(\delta) \geq \varepsilon) = 0,
\]
where
\[
\omega''(\delta) = \sup \{|\phi^x(u) - \phi^x(u_1)| \wedge |\phi^x(u_2) - \phi^x(u)| : u_1 \leq u \leq u_2, u_2 - u_1 \leq \delta\}.
\]

From the mean value theorem we know that there exists a constant \( C \) such that for \( u < v \in [a, b] \)
\[
|\phi^x(u) - \phi^x(v)| \leq C(v - u)|T_x - \frac{1}{r} \log x| \left( e^{\lambda(v) |T_x - \frac{1}{r} \log x|} + e^{r(v) |T_x - \frac{1}{r} \log x|} \right)
\leq 2C(v - u)|T_x - \frac{1}{r} \log x| e^{\lambda(v) |T_x - \frac{1}{r} \log x|}
\leq 2C(v - u)e^{(\lambda + 1)|T_x - \frac{1}{r} \log x|},
\]
where \( \lambda \geq \max(\lambda_1, r) \). Thus if \( \omega''(\delta) \geq \varepsilon \) then necessarily
\[
2C\delta e^{(\lambda + 1)|T_x - \frac{1}{r} \log x|} \geq \varepsilon,
\]
and therefore
\[
\mathbb{P}(\omega''(\delta) \geq \varepsilon) \leq \mathbb{P} \left( |T_x - \frac{1}{r} \log x| \geq \frac{b}{\lambda + 1} \log(\varepsilon/\delta) \right).
\]
The condition (3) then follows by taking the limit as \( x \to \infty \) (using (1)) and then sending \( \delta \) to 0. 

4.1 Crossover time

In this section we investigate, using the limit theorems proven in the previous section, the distribution of the time:
\[
\xi \equiv \inf \{ t \in (0, 1] | Z_1(t) \geq Z_0(t) \}.
\]
We will refer to this time 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. First we utilize the crossover time of the limit processes to estimate $\xi$. In particular, we define $\tilde{u}$ to be the solution to:

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

Setting these equal we obtain

$$e^{\lambda_0 \tilde{u}(T_x - \frac{1}{r} \log x)x^{1-\tilde{u}}} = \frac{r_0 \mu}{\lambda_1 - \lambda_0}x^{1-\alpha}e^{\lambda_1 \tilde{u}T_x} \left(1 - e^{(\lambda_0 - \lambda_1)\tilde{u}T_x}\right)$$

and solving:

$$\tilde{u} = \log \left(\frac{r_0 \mu + (\lambda_1 + r)x^\alpha}{xT_x(\lambda_1 + r)} - \log r_0 \mu\right).$$

Next we prove the following result:

**Theorem 5** The estimate $\tilde{u}$ converges to the true crossover time, $\xi$, in probability as $x \to \infty$, i.e., for all $\varepsilon > 0$

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

The proof of this result follows as an application of Theorem 3, see Section 7 for the proof.

### 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 random time associated with the unscaled escape process, which we will call the turnaround time:

$$\tau = \arg\min_{t \geq 0} \{Z_0(t) + Z_1(t)\}.$$ 

We will approximate the turnaround time using the sample path approximations, as the random variable

$$u^* \equiv \frac{\log (\frac{r_0 \mu}{\lambda_1 + r}) + \log (x^\alpha(\lambda_1 + r) - r_0 \mu)}{(\lambda_1 + r)T_x} \tag{4}$$

which 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), we see that for large $x$ we get the following approximation

$$u^* \approx \frac{\alpha r}{\lambda_1 + r}. \tag{5}$$

Therefore, it follows that higher mutation rates, or smaller $\alpha$, leads to a quicker turnaround time (relative to the extinction time). In addition, we can see that as we increase the decay rate $r$, the time of progression relative to the time of extinction increases. Throughout this section we
will 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
\[
f_{x,t}(u) = EZ_0(us_x(t)) + EZ_1(us_x(t))
\]
\[
= xe^{\lambda_0 u(1/2 \log x + t)} \left( 1 - \frac{r_0 \mu}{x^{\alpha}(1 + r)} \right) \left( 1 + \frac{r}{\lambda_1} \right)
\]
We will first show that the following critical point of \( f_{x,t} \)
\[
u^*(t) \equiv \frac{\log \left( \frac{r}{\lambda_1 + r} \right) + \log \left( x^{\alpha}(1 + r) - r_0 \mu \right)}{(1/2 \log x + t)}
\]  
(6)
is with high probability the minimum of \( Z_0(us_x(t)) + Z_1(us_x(t)) \). We will then show that this statement is in fact true uniformly for \( t \) in compact sets, and with this we will be able to show that \( \tau/T_x \) 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
\[
f_{x,t}(u^*(t)) = xe^{\lambda_0 u^*(t)(1/2 \log x + t)} \left( 1 - \frac{r_0 \mu}{x^{\alpha}(1 + r)} \right) \left( 1 + \frac{r}{\lambda_1} \right),
\]
and
\[
f_{x,t}(u^*(t) + y) = xe^{\lambda_0 u^*(t)(1/2 \log x + t)} \left( 1 - \frac{r_0 \mu}{x^{\alpha}(1 + r)} \right) \left( e^{\lambda_0 y(1/2 \log x + t)} + \frac{r}{\lambda_1} e^{\lambda_1 y(1/2 \log x + t)} \right).
\]
Therefore
\[
f_{x,t}(u^*(t) + y) = f_{x,t}(u^*(t)) \left[ \left( \frac{\lambda_1}{\lambda_1 + r} \right) \left( x^{-y} e^{\lambda_0 yt} + \frac{r}{\lambda_1} e^{\lambda_1 yt} \right) \right].
\]  
(7)

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 \( u^*(t) \). In particular we can establish the following result,

**Lemma 3** For \( \varepsilon > 0 \),
\[
P \left( \left| u^*(t) - \frac{\tau}{1/2 \log x + t} \right| > \varepsilon \right) \to 0
\] as \( x \to \infty \).

Similar to the approximation theorem, once the result is established for arbitrary \( t \), it is possible to show that the result hold uniformly for \( t \) in compact sets.

**Lemma 4** For \( \varepsilon > 0 \) and a constant \( M > 0 \),
\[
P \left( \sup_{t \in [-M,M]} \left| u^*(t) - \frac{\tau}{1/2 \log x + t} \right| > \varepsilon \right) \to 0.
\] as \( x \to \infty \).
Figure 2: Sample PDF of $\tau/T_x$ 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: $x = 100000, r_0 = 1.0, d_0 = 1.75, r_1 = 2.0, d_1 = 1.0, \mu = 0.01$.

See Section 7 for details of the proofs of the two lemmas above.

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

**Theorem 6** For $\varepsilon > 0$,

$$\mathbb{P}\left(\left|u^* - \frac{\tau}{T_x}\right| > \varepsilon\right) \to 0.$$  

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 $\tau/T_x$ 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.
Figure 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]).

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. 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 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 $\tau$ is a clinically observable quantity since it represents the time until disease progression from the start of treatment. Once $\tau$ is observed, using
Theorem 6 and the approximation in (5) 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 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 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. Lastly, we mention the Markovian assumption of our model is only used in the integral representation for the conditional distribution of the resistant population; thus extensions to the non-Markovian setting seem possible. 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 will 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 that

$$\mathbb{E}[Z_1(vT_x)] = r_0 \mu \mathbb{E} \left[ \int_0^{vT_x} Z_0(y) \exp(\lambda_1(vT_x - y)) \, dy \right].$$

We can then rearrange the order of integration (by Tonelli’s Thm) to see

$$\mathbb{E}[Z_1(vT_x)] = r_0 \mu \mathbb{E} \left[ \int_0^{\infty} 1_{T_x \geq y/v} Z_0(y) \exp(\lambda_1(vT_x - y)) \, dy \right]$$

$$= r_0 \mu \int_0^{\infty} \mathbb{E} \left[ 1_{T_x \geq y/v} Z_0(y) \exp(\lambda_1(vT_x - y)) \right] \, dy$$

$$= r_0 \mu \int_0^{\infty} \int_{y/v}^{\infty} \sum_{n=1}^{\infty} ne^{\lambda_1(vt-y)} P(T_x \in dt | Z_0(y) = n) P(Z_0(y) = n) \, dy.$$
Note that \( P(T_x \in dt | Z_0(y) = n) = g_n(t - y) dt \), 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 \( g_n(t) = n (G(t))^{n-1} g(t) \), 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

\[
E[Z_1(vT_x)] = r_0 \mu \int_0^\infty \int_{y/v}^\infty e^{\lambda_1(vt-y)} g(t-y) \left( \sum_{n=1}^{\infty} n^2 (G(t-y))^{n-1} P(Z_0(y) = n) \right) dt dy.
\]

Next define

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

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

\[
\sum_{n=1}^{\infty} n^2 s^{n-1} P(Z_0(t) = n) = s \frac{\partial^2}{\partial s^2} F_x(s, t) + \frac{\partial}{\partial s} F_x(s, t)
\]

\[
= sx (F(s, t))^{x-1} \frac{\partial^2}{\partial s^2} F(s, t) + sx (x-1) (F(s, t))^{x-2} \left( \frac{\partial}{\partial s} F(s, t) \right)^2
\]

\[
+ x (F(s, t))^{x-1} \frac{\partial}{\partial s} F(s, t).
\]

For ease of notation we will simply write \( \frac{\partial}{\partial s} F(s, t) = F'(s, t) \). Plugging the formula from display (8) into the expression for the mean we get

\[
E[Z_1(vT_x)] = x r_0 \mu \int_0^\infty \int_{y/v}^\infty e^{\lambda_1(vt-y)} g(t-y) (F(G(t-y), y))^{x-2} \left[ (x-1)G(t-y) \left( F'(G(t-y), y) \right)^2 + G(t-y)F'(G(t-y), y)F''(G(t-y), y) + F(G(t-y), y)F'(G(t-y), y) \right] dy dt.
\]

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

\[
t = \frac{1}{r} \left( z + \log \frac{x}{b} \right) = z_x,
\]

and observe that

\[
e^{\lambda_1(vz_x-y)} = e^{\lambda_1(vz/r-y)} \left( \frac{x}{b} \right)^{\lambda_1 v/r}.
\]

After the change of variables we have

\[
E[Z_1(vT_x)] = \frac{x}{b} \frac{r_0 \mu}{b \lambda_{1v}} \int_0^\infty \int_{v \log z_x}^\infty e^{\lambda_1(vz/r-y)} g(z_x - y) G(z_x - y) (F(G(z_x - y), y))^{x-2} \left( F'(G(z_x - y), y) \right)^2 dz dy
\]

\[
+ \frac{x}{b} \frac{r_0 \mu}{b \lambda_{1v}} \int_0^\infty \int_{v \log z_x}^\infty e^{\lambda_1(vz/r-y)} g(z_x - y) (F(G(z_x - y), y))^{x-1} [G(z_x - y) F''(G(z_x - y), y) + F'(G(z_x - y), y)] dz dy
\]

\[
= I_1(x, v) + I_2(x, v).
\]

18
Following an argument similar to Jagers et al 2007, we will now establish that for \( v \in (0, 1] \), \( x^{-1-\frac{\lambda_1 v}{2}} I_1(x, v) = \tilde{I}_1(x, v) \to I_1(v) \) and \( x^{-1-\frac{\lambda_1 v}{2}} I_2(x, v) = \tilde{I}_2(x, v) \to 0 \) as \( x \to \infty \).

The integrand of \( \tilde{I}_1(x, v) \) is

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

where \( c_0 = r_0 \mu b^{-\lambda_1 v / r} \). From \([10]\) we know that as \( z \to \infty \), \( g(z) \sim re^{-r z} / b \), and therefore

\[
g(z - y) \sim \frac{r}{x} e^{r z} e^{-z}, \tag{9}
\]
as \( x \to \infty \). Next notice that there exists a \( \xi_x \in (G(z_x - y), 1) \) such that

\[
F'(G(z_x - y), y) = F'(1, y) + (1 - G(z_x - y)) F''(\xi_x, y) = e^{-r y} + O(1 - G(z_x - y)),
\]

and therefore

\[
F'(G(z_x - y), y) \sim e^{-r y}. \tag{10}
\]

Lastly observe that

\[
F(G(z_x - y), y) = 1 + e^{-r y} (G(z_x - y) - 1) + O(G(z_x - y) - 1)^2,
\]

and therefore

\[
\log F(G(z_x - y), y) = -(1 - G(z_x - y)) e^{-r y} + O(G(z_x - y) - 1)^2.
\]

Observe that

\[
1 - G(z_x - y) \sim \frac{e^{r y} e^{-z}}{x},
\]

which gives that

\[
(F(G(z_x - y), y))^{x-2} \sim \exp[-e^{-z}]. \tag{11}
\]

Combining (9), (10), and (11) we see that

\[
\lim_{x \to \infty} f_x(z, y) = c_1 e^{-z(1-\frac{\lambda_1 v}{2})} e^{-y(r+\lambda_1)} \exp[-e^{-z}] = f(z, y), \tag{12}
\]

where \( c_1 = r c_0 \). 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 \( f_x(z, y) \leq h(z, y) \). First note that since \( G(z) \leq 1 \), and \( F(s, t) \leq E_1 Z_0(t) = e^{-r t} \) we have

\[
f_x(z, y) \leq c_0 x e^{\lambda_1 (z^2 / r - y)} g(z - y) (F(G(z_x - y), y))^{x-2} e^{-2r y}.
\]

Then observe that there exists a constant \( k_1 \) such that

\[
g(z - y) \leq k_1 e^{-r(z_x - y)} = \frac{k_2}{x} e^{-z} e^{-r y}. \tag{13}
\]
Since $\log x \leq x - 1$ we have
\[
(F(G(z_x - y), y))^x - 2 = \exp [(x - 2) \log F(G(z_x - y), y)] \leq \exp [-(x - 2)(1 - F(G(z_x - y), y))].
\]
Due to the convexity in $s$ of $F(s, t)$ we have that $F(s, t) = 1 + F'(1, t)(s - 1) + (s - 1)^2 F''(\xi, t)/2 \geq 1 + F'(1, t)(s - 1)$, and therefore
\[
(F(G(z_x - y), y))^{x - 2} \leq \exp [-(x - 2)(1 - G(z_x - y))e^{-y}] .
\]
Then we can find a constant $k_2$ such that $1 - G(z) \geq k_2 e^{-rz}$, and we arrive at
\[
(F(G(z_x - y), y))^{x - 2} \leq \exp [-(x - 2)e^{-ry}k_2e^{-r(z_x - y)}] \leq \exp \left[ -\frac{k_2}{2} e^{-z} \right] . \tag{14}
\]
Based on (13) and (14) we see that we can use the dominating function
\[
h(z, y) = k_3 e^{\lambda_1(\frac{u}{r} - y)} e^{-z} e^{-ry} \exp [k_2 e^{-z}].
\]
With this result we see that (need to check constants)
\[
\lim_{x \to \infty} \tilde{I}_1(x, v) = \frac{r_0 \mu}{r} \int_0^\infty \int_{-\infty}^\infty \lim_{x \to \infty} f_x(z, y) dz dy
\]
\[
= \frac{r_0 \mu}{r} \int_0^\infty \int_{-\infty}^\infty e^{\lambda_1(\frac{u}{r} - y)} e^{-ryc^*} e^{-z} \exp [k_2 e^{-z}] dz dy
\]
\[
= \frac{r_0 \mu \Gamma(1 - \frac{\lambda_1}{r})}{b^{\lambda_1 v/r}(\lambda_1 + r)}.
\]

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
\[
\tilde{I}_2(x, v) \leq k_5 x^{-v\lambda_1/r} \int_0^\infty \int_{y/v}^\infty e^{\lambda_1(ut - y)} g(t - y) dt dy
\]
\[
= k_5 x^{-v\lambda_1/r} \int_0^\infty \int_{y/(1 - v)}^\infty e^{-\lambda_1 y} g(s) e^{\lambda_1 v(s + y)} ds dy = k_6 x^{-v\lambda_1/r},
\]
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

**Proof.** We will prove the more difficult second statement first. We observe that it suffices to prove the following converges to 0 as $x \to \infty$,
\[
\mathbb{P} \left( \sup_{u \in [0, 1]} e^{-\lambda_1 ut} |Z^x_1(us_x(t)) - \phi^x_1(u, t)| > e^{-\lambda_1 t x} \right) = 0.
\]

20
Next observe that

\[
\frac{e^{-\lambda_1 ut}}{x^{-\alpha + 1 + \lambda_1 u/r}} (Z_1(us_x(t)) - \phi_1(u, t)) = \left( \frac{e^{-\lambda_1 ut}}{x^{1+\lambda_1 u/r}} Z_1(us_x(t)) - \frac{r_0\mu}{x^{1+\alpha}} \int_0^{us_x(t)} Z_0(s)e^{-\lambda_1 s} ds \right) x^\alpha + \frac{r_0\mu}{x} \int_0^{us_x(t)} (Z_0(s) - xe^{\lambda_0 s}) e^{-\lambda_1 s} ds,
\]

and therefore via the triangle inequality we have

\[
P \left( \sup_{u \in [0, 1]} e^{-\lambda_1 ut} |Z_1^x(us_x(t)) - \phi_1^x(u, t)| > e^{-\lambda_1 t} \varepsilon \right) \leq P \left( \sup_{u \in [0, 1]} x^\alpha \left| \frac{e^{-\lambda_1 ut}}{x^{1+\lambda_1 u/r}} Z_1(us_x(t)) - \frac{r_0\mu}{x^{1+\alpha}} \int_0^{us_x(t)} Z_0(s)e^{-\lambda_1 s} ds \right| > \varepsilon/2 \right)
\]

\[
+ P \left( \sup_{u \in [0, 1]} \frac{r_0\mu}{x} \int_0^{us_x(t)} |Z_0(s) - xe^{\lambda_0 s}| e^{-\lambda_1 s} ds > \varepsilon/2 \right).
\]

However we can observe that the process considered in final expression is monotonic in \( u \), and the process considered in the second expression is a martingale in \( u \), which allows us to arrive at the following simpler inequality,

\[
P \left( \sup_{u \in [0, 1]} e^{-\lambda_1 ut} |Z_1^x(us_x(t)) - \phi_1^x(u, t)| > e^{-\lambda_1 t} \varepsilon \right) \leq \frac{4x^{2\alpha}}{2\varepsilon^2} \mathbb{E} \left[ \left( \frac{e^{-\lambda_1 t}}{x^{1+\lambda_1 u/r}} Z_1(s_x(t)) - \frac{r_0\mu}{x^{1+\alpha}} \int_0^{s_x(t)} Z_0(s)e^{-\lambda_1 s} ds \right)^2 \right]
\]

\[
+ P \left( \frac{r_0\mu}{x} \int_0^{s_x(t)} |Z_0(s) - xe^{\lambda_0 s}| e^{-\lambda_1 s} ds > \varepsilon/2 \right).
\]  

(15)

Consider the latter quantity first, where it of course suffices to show that the following converges to 0 with \( x \),

\[
\frac{r_0\mu}{x} \int_0^{s_x(t)} \mathbb{E} \left[ \left| Z_0(s) - xe^{\lambda_0 s} \right| e^{-\lambda_1 s} ds \right].
\]

We then have the following from the Cauchy-Schwartz inequality,

\[
\mathbb{E} \left[ \left| Z_0(s) - xe^{\lambda_0 s} \right| \right] \leq (\text{Var}(Z_0(s)))^{1/2}.
\]

Next observe that,

\[
\text{Var}(Z_0(s)) = x \left( \frac{r_0 + d_0}{\lambda_0} \right) \left( e^{2\lambda_0 t} - e^{\lambda_0 t} \right),
\]
and it follows that
\[
\frac{r_0 \mu}{x} \int_0^{s_x(t)} E \left[ \left( Z_0(s) - xe^{\lambda_0 s} \right) e^{-\lambda_1 s} ds \right] = O(x^{-1/2}).
\]

Moving onto the first term in (15), we expand that term to get
\[
x^{2\alpha} E \left[ \left( \frac{e^{-\lambda_1 t}}{x^{1+\alpha}} \right)^2 \int_0^{s_x(t)} Z_0(s)e^{-\lambda_1 s} ds \right] = \left( \frac{e^{-\lambda_1 t} x^{\alpha}}{x^{1+\lambda_1/\alpha}} \right)^2 E \left[ Z_1(s_x(t))^2 \right]
\]
\[
- 2 x^{2\alpha} r_0 \mu e^{-\lambda_1 t} \int_0^{s_x(t)} e^{-\lambda_1 s} E \left[ Z_0(s)Z_1(s_x(t)) \right] ds
\]
\[
+ \left( \frac{r_0 \mu}{x} \right)^2 \int_0^{s_x(t)} \int_0^{s_x(t)} E[Z_0(s)Z_0(y)]e^{-\lambda_1 s}e^{-\lambda_1 y} ds dy.
\]

Assume for now that
\[
E \left[ Z_1(s_x(t))^2 \right] = \left( \frac{r_0 \mu}{x^\alpha} \right)^2 \int_0^{s_x(t)} \int_0^{s_x(t)} E[Z_0(s)Z_0(y)]e^{\lambda_1(s_x(t)-s)}e^{\lambda_1(s_x(t)-s)} ds dy
\]
\[
+ \frac{r_0 \mu}{x^\alpha} \int_0^{s_x(t)} E[Z_0(s) \left( 2r_1 e^{2\lambda_1(s_x(t)-s)} - (r_1 + d_1) e^{\lambda_1(s_x(t)-s)} \right) ds
\]
and
\[
E \left[ Z_0(s)Z_1(s_x(t)) \right] = e^{\lambda_1 s_x(t)} \frac{r_0 \mu}{x^\alpha} \int_0^{s_x(t)} E[Z_0(y)Z_0(s)]e^{-\lambda_1 y} dy.
\]

Then plugging the expressions in (17) and (18) into the expression in (16) we easily see that
\[
x^{2\alpha} E \left[ \left( \frac{e^{-\lambda_1 t}}{x^{1+\alpha}} \right)^2 \int_0^{s_x(t)} Z_0(s)e^{-\lambda_1 s} ds \right] = \frac{2r_1 x^{\alpha-1} r_0 \mu}{\lambda_1} \int_0^{s_x(t)} e^{s(\lambda_0-2\lambda_1)} ds - \frac{(r_1 + d_1) r_0 \mu e^{-\lambda_1 t}}{\lambda_1 x^{1+\alpha+\lambda_1/\alpha}} \int_0^{s_x(t)} e^{s(\lambda_0-\lambda_1)} ds
\]
\[
= O \left( x^{\alpha-1} \right).
\]

Therefore the proof of the lemma will be completed once we establish the equalities in (17) and (18).

We will start by establishing the latter result. Define \( \mathcal{F}_0^\infty \) to be the sigma algebra generated by the wave 0 population until their eventual extinction, then
\[
E \left[ Z_1(s_x(t)) \big| \mathcal{F}_0^\infty \right] = r_0 \mu_x \int_0^{s_x(t)} Z_0(y)e^{\lambda_1(s_x(t)-y)} dy
\]
and therefore
\[
E[Z_0(s)Z_1(s_x(t))] = e^{\lambda_1s_x(t)}r_0\mu_x \int_0^{s_x(t)} E[Z_0(y)Z_0(s)]e^{-\lambda_1y}dy
\]

Now we establish the result in (17). For simplicity we evaluate \(E[Z_1(t)]\) for a positive \(t\). For ease of notation we will use the following \(\tilde{E}[] = E[\cdot|F_\infty^0]\). 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

\[
E[Z_1(t)^2|F_\infty^0] = \tilde{E}\left(\sum_{j=0}^{m} \sum_{k=1}^{N_j} B_{j,k}(t-\tau_{j,k})\right)^2
\]

\[
= \sum_{j=0}^{m} \tilde{E}\left(\sum_{k=1}^{N_j} B_{j,k}(t-\tau_{j,k})\right)^2
\]

\[
+ \sum_{j=0}^{m} \sum_{\ell \neq j} \tilde{E}\left(\sum_{k=1}^{N_j} B_{j,k}(t-\tau_{j,k})\right) \tilde{E}\left(\sum_{k=1}^{N_\ell} B_{\ell,k}(t-\tau_{\ell,k})\right),
\]

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. One can easily calculate that for \(0 \leq j \leq m\)

\[
E[N_j] = \Delta r_0\mu_x Z_0(j\Delta) + o(\Delta)
\]

\[
E[N_j^2] = \Delta r_0\mu_x Z_0(j\Delta)(1 + \Delta r_0\mu_x Z_0(j\Delta)) + o(\Delta).
\]

Therefore

\[
\tilde{E}\left(\sum_{k=1}^{N_j} B_{j,k}(t-\tau_{j,k})\right)^2 = \tilde{E}N_j \tilde{E}B(t-\tau_j)^2 + \tilde{E}N_j^2 (\tilde{E}B(t-\tau_j))^2
\]

\[
= \Delta r_0\mu_x Z_0(j\Delta) \tilde{E}B(t-\tau_j)^2 + O(\Delta^2)
\]

\[
= \Delta r_0\mu_x Z_0(j\Delta) \left(\frac{2r_1}{\lambda_1} e^{2\lambda_1(t-\tau_j)} - \frac{r_1 + d_1}{\lambda_1} e^{\lambda_1(t-\tau_j)}\right) + O(\Delta^2),
\]

and

\[
\tilde{E}\left(\sum_{k=1}^{N_\ell} B_{\ell,k}(t-\tau_{\ell,k})\right) = \Delta r_0\mu_x Z_0(j\Delta) \tilde{E}B(t-\tau_j)
\]

\[
= \Delta r_0\mu_x Z_0(j\Delta) e^{\lambda_1(t-\tau_j)}.
\]
Using the previous two expressions we get

\[
\tilde{E}[Z_1(t)^2] = \Delta r_0 \mu_x \sum_{j=0}^m Z_0(j \Delta) \left( \frac{2r_1}{\lambda_1} 2^{\lambda_1 (t-\tau_j)} - \frac{r_1 + d_1}{\lambda_1} e^{\lambda_1 (t-\tau_j)} \right) \\
+ (r_0 \mu_x \Delta)^2 \sum_{j=0}^m \sum_{\ell=0, \ell \neq j}^m Z_0(j \Delta) Z_0(\ell \Delta) e^{\lambda_1 (t-\tau_j)} e^{\lambda_1 (t-\tau_\ell)}.
\]

Sending \( \Delta \to 0 \), integrating over \( Z_0 \), and replacing \( t \) with \( \frac{1}{\theta} \log x + t \) gives us the desired formula in (17).

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

\[
\sup_{u \in [0,a]} |Z^x_0(us_x(t)) - \phi^x_0(u,t)| \\
\leq e^{-\lambda_0 at} \sup_{u \in [0,a]} e^{-\lambda_0 at} |Z^x_0(us_x(t)) - \phi^x (u,t)|.
\]

Therefore we will show that as \( x \to \infty \) the following converges to 0,

\[
P \left( \sup_{u \in [0,a]} |Z^x_0(us_x(t)) - \phi^x_0(u,t)| \geq \varepsilon e^{-\lambda_0 at} \right) \\
= P \left( \sup_{u \in [0,a]} \left| e^{-\lambda_0 at} Z^x_0(us_x(t)) - 1 \right| \geq \varepsilon e^{-\lambda_0 at} \right).
\]

Next observe that, \( e^{-\lambda_0 at} x^{u-1} Z_0(us_x(t)) - 1 \) is a martingale with respect to \( u \) and therefore it suffices to show the following converges to 0 as \( x \to \infty \)

\[
E \left[ e^{-\lambda_0 at} Z^x_0(as_x(t)) - 1 \right]^2 \\
= E \left[ \left( e^{-\lambda_0 at} Z^x_0(as_x(t)) \right)^2 \right] - 1 \\
= x^{2a-2} e^{-2\lambda_0 at} \left( x \text{Var}_1 Z_0(as_x(t)) + x^2 e^{2\lambda_0 as_x(t)} \right) - 1 \\
= x^{2a-2} e^{-2\lambda_0 at} \text{Var}_1 Z_0(as_x(t)),
\]

where \( \text{Var}_1 Z_0(t) \) represents the variance of \( Z_0(t) \) starting with an initial population size 1. Using the formula for the variance of the binary branching process again, we see that the previous expression reduces to,

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

and since we have assumed that \( a < 1 \), the result is established. \( \blacksquare \)
7.3 Proof of Lemma 2

**Proof.** Throughout the proof assume that \( x > e^{rM} \). We first establish the result for the \( Z_1 \) population by showing a monotonicity property in \( t \). In particular, we will now establish that

\[
\sup_{u \in [0,1]} \left| Z_1^x \left( u \left( \frac{1}{r} \log x + t_0 \right) \right) - \phi_1^x (u, t_0) \right| \leq \sup_{u \in [0,1]} \left| \frac{1}{r} \log x + t_1 \right| - \phi_1^x (u, t_1) \tag{19}
\]

for \( t_0 \leq t_1 \). For any \( u \in [0,1] \) set \( \bar{u} = u \left( \frac{1}{r} \log x + t_1 \right) \), 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 \( \bar{u} \leq u \) and thus \( x^{-\lambda_1 u/r} \leq x^{-\lambda_1 \bar{u}/r} \). Therefore,

\[
x^{-\lambda_1 u/r} \left| Z_1 \left( u \left( \frac{1}{r} \log x + t_0 \right) \right) - \phi_1 (u, t_0) \right| \leq x^{-\lambda_1 \bar{u}/r} \left| Z_1 \left( \bar{u} \left( \frac{1}{r} \log x + t_1 \right) \right) - \phi_1 (\bar{u}, t_1) \right|.
\]

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

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

Result (ii) now follows by an application of Lemma 1.

The proof of result (i) will follow a similar approach. In particular for \( t \leq M \), and \( u \in [0,a] \) define \( \tilde{u} = u \left( \frac{1}{r} \log x + t \right) \).

Notice that

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

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

\[
x^u \left| Z_0 (u s_x(t)) - \phi_0 (u, t) \right|
\]

\[
eq x^{u - \tilde{u}} \tilde{u}^u Z_0 \left( \tilde{u} \left( \frac{1}{r} \log x + M \right) \right) - \phi_0 (\tilde{u}, M)
\]

\[
\leq e^{2arM} x^{u - \tilde{u}} Z_0 \left( \tilde{u} \left( \frac{1}{r} \log x + M \right) \right) - \phi_0 (\tilde{u}, M).
\]

25
Since the previous inequality holds for any $u$, we know that for any $t \in [-M, M]$, 
\[
\sup_{u \in [0,a]} x^u |Z_0(us_x(t)) - \phi_0(u,t)| 
\leq e^{2arM} \sup_{u \in [0,a]} x^u \left| Z_0 \left( u \left( \frac{1}{r} \log x + M \right) \right) - \phi_0(u,M) \right|.
\]
Thus the result of (i) is established by using the result of Lemma 1 for $t = M$. ■

### 7.4 Proof of Theorem 5

**Proof.** First we prove that 
\[
P(\xi \leq \tilde{u} - \epsilon) \to 0
\]
as $x \to \infty$. In particular recall that 
\[
\phi_0(u,t) = x^{1-u}e^{\lambda_0 ut} \\
\phi_1(u,t) = \frac{r_0 x^{1-\alpha + \lambda_1 u/r} e^{\lambda_1 ut}}{\lambda_1 + r} \left( 1 - e^{u(\lambda_0 - \lambda_1) t_x} e^{(\lambda_0 - \lambda_1) u/r} \right).
\]
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:

\[
P \left( \sup_{u \leq \tilde{u} - \epsilon} (Z_1(uT_x) - Z_0(uT_x)) > 0 \right)
\]
\[
= P \left( \sup_{u \leq \tilde{u} - \epsilon} x^{u-1}(Z_1(uT_x) - \phi_1(u, d(T_x))) + (\phi_1(u,d(T_x)) - \phi_0(u,d(T_x)))
+ (\phi_0(u,d(T_x)) - Z_0^x(uT_x)) > 0 \right)
\]
\[
\leq P \left( \sup_{u \leq \tilde{u} - \epsilon} x^{u-1}(Z_1(uT_x) - \phi_1(u,d(T_x))) + \sup_{u \leq \tilde{u} - \epsilon} x^{u-1}(\phi_1(u,d(T_x)) - \phi_0(u,d(T_x))) +
+ \sup_{u \leq \tilde{u} - \epsilon} x^{u-1}(\phi_0(u,d(T_x)) - Z_0^x(uT_x)) > 0 \right) \quad (20)
\]
To study this let us start by considering the first term in the sum above:

\[
\sup_{u \leq \tilde{u} - \epsilon} \left| x^{u-1}(Z_1(uT_x) - \phi_1(u,d(T_x))) \right| \quad (21)
\]
\[
\leq \sup_{u \leq \tilde{u} - \epsilon} x^{u(1 + \lambda_1) - \alpha} \sup_{u \leq \tilde{u} - \epsilon} \left| x^{u-\alpha} \left( Z_1(uT_x) - \phi_1(u,d(T_x)) \right) \right|.
\]
The second term in the product converges to zero in probability via Theorem 3. The first term tends to zero by the following argument:
\[
\log \left[ \sup_{u \leq \tilde{u} - \epsilon} x^{u(1 + \frac{\lambda_1}{r}) - a} \right] = \log \left[ x^{(\tilde{u} - \epsilon)(1 + \frac{\lambda_1}{r}) - a} \right] \\
= \log \left[ x^{-(1 + \frac{\lambda_1}{r})\epsilon - a} \exp[\tilde{u}(1 + \frac{\lambda_1}{r}) \log x] \right] \\
= \left( -\alpha - (1 + \frac{\lambda_1}{r})\epsilon \right) \log x + \frac{\log x}{r T_x} \log \left[ 1 + x^\alpha \left( \frac{\lambda_1 + r}{r_0 \mu} \right) \right] \\
\leq \left( -\alpha - (1 + \frac{\lambda_1}{r})\epsilon \right) \log x + \frac{\log x}{r T_x} \log \left[ 2x^\alpha \left( \frac{\lambda_1 + r}{r_0 \mu} \right) \right] \\
= \alpha \frac{\log x}{r T_x} (\log x - r T_x) - \left( \frac{\lambda_1}{r} + 1 \right) \epsilon \log x + \frac{\log x}{r T_x} \log \left[ \frac{2(\lambda_1 + 1)}{r_0 \mu} \right],
\]

where in the third equality we have utilized the fact that
\[
\tilde{u} \left( \frac{\lambda_1 + r}{r} \right) = \frac{1}{r T_x} \log \left[ \frac{(\lambda_1 + r)x^\alpha}{r_0 \mu} + 1 \right]
\]
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
\[
\frac{\log x}{r T_x} \rightarrow 1
\]
in probability. Thus, we conclude that (21) goes to zero in probability.

The second term in (20) is analyzed next.
\[
\sup_{u \leq \tilde{u} - \epsilon} x^{u-1}(\phi_1(u, d(T_x)) - \phi_0(u, d(T_x))) = x^{\tilde{u}-\epsilon-1}(\phi_1(\tilde{u} - \epsilon, d(T_x)) - \phi_0(\tilde{u} - \epsilon, d(T_x)))
\]
\[
= x^{\tilde{u}-\epsilon-1} \phi_1(\tilde{u} - \epsilon, d(T_x)) - x^{\tilde{u}-\epsilon-1} \phi_0(\tilde{u} - \epsilon, d(T_x)).
\]

Via the definition of \( \phi_0(u, d(T_x)) \) and the limit result on the extinction time (1), we have that as \( x \rightarrow \infty \)
\[
x^{\tilde{u}-\epsilon-1} \phi_0(\tilde{u} - \epsilon, d(T_x)) \Rightarrow c^{-(\tilde{u}-\epsilon)}e^{-(\tilde{u}-\epsilon)\eta},
\]
where \( \eta \) 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
\[
x^{\tilde{u}-\epsilon-1} \phi_1(\tilde{u} - \epsilon, d(T_x)) = x^{(\tilde{u}-\epsilon)(1 + \frac{\lambda_1}{r}) - a \alpha(\tilde{u}-\epsilon)\lambda_1/\lambda_1 - 1} \phi_1(\tilde{u} - \epsilon, d(T_x))
\]
where the first term approaches zero, as argued previously since its log approaches negative infinity, and the product of the remaining terms approaches a constant times the exponential of a Gumbel, which is again a result of (1). The third term in (20)

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

converges to zero in probability by Theorem 3. Therefore

$$\limsup_{x \to \infty} \mathbb{P}(\xi \leq \tilde{u} - \epsilon) \leq \mathbb{P}(e^{-(\tilde{u} - \epsilon)}e^{-(\tilde{u} - \epsilon)\eta} \leq 0) = 0.$$ 

Next, we need to show that

$$\mathbb{P}(\xi \geq \tilde{u} + \epsilon) \to 0.$$ 

We have by definition of $\xi$, that

$$\mathbb{P}(\xi > \tilde{u} + \epsilon) \leq \mathbb{P}(Z_0((\tilde{u} + \epsilon)T_x) - Z_1((\tilde{u} + \epsilon)T_x) > 0)$$

$$= \mathbb{P}(x^{\alpha-(\tilde{u}+\epsilon)} \lambda_1^{\frac{1}{r} - 1}(Z_0((\tilde{u} + \epsilon)T_x) - Z_1((\tilde{u} + \epsilon)T_x) > 0)$$

$$+ x^{\alpha-(\tilde{u}+\epsilon)} \lambda_1^{\frac{1}{r} - 1}(\phi_0(\tilde{u} + \epsilon, d(T_x)) - \phi_1(\tilde{u} + \epsilon, d(T_x)))$$

$$+ x^{\alpha-(\tilde{u}+\epsilon)} \lambda_1^{\frac{1}{r} - 1}(\phi_1(\tilde{u} + \epsilon, d(T_x)) - Z_1((\tilde{u} + \epsilon)T_x) > 0)$$

(22)

By Theorem 3 the third term goes to zero. We observe that the first term can be written as

$$x^{\alpha-(\tilde{u}+\epsilon)} \lambda_1^{\frac{1}{r} + 1} x^{\tilde{u} + \epsilon - 1}(Z_0((\tilde{u} + \epsilon)T_x) - \phi_0(\tilde{u} + \epsilon, d(T_x))).$$

Then, arguing in a similar fashion as previously,

$$\log \left[ x^{\alpha-(\tilde{u}+\epsilon)} \lambda_1^{\frac{1}{r} + 1} \right] = \left( \alpha - \epsilon \left( \frac{\lambda_1}{r} + 1 \right) \right) \log x - \frac{\log x}{rT_x} \log \left( \frac{\lambda_1 + r}{r_0\mu} \right) x^\alpha + 1$$

$$\leq -\alpha \frac{\log x}{rT_x} \log x - \frac{\lambda_1}{r} + 1 \epsilon \log x - \frac{\log x}{rT_x} \log \left[ \frac{2(\lambda_1 + 1)}{r_0\mu} \right],$$

is seen to diverge to $-\infty$. Thus, $x^{\alpha-(\tilde{u}+\epsilon)} \lambda_1^{\frac{1}{r} + 1}$ approaches zero as $x \to \infty$. This fact allows us to conclude that the first term in (22) approaches zero. Similar to the previous argument, the second term in (22) approaches a limiting random variable, which is characterized by the weak limit of

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

Since this is a strictly negative random variable, we establish the result. \hfill \blacksquare
7.5 Proof of Lemma 3

Here we establish Lemma 3, namely that $u^*(t)$ approximates $\tau/(\frac{1}{r} \log x + t)$.

**Proof.** We will prove first that

$$\mathbb{P} \left( \inf_{u \in [u^*(t) + \epsilon, \infty]} Z_1(us_x(t)) < Z_0(u^*(t) s_x(t)) + Z_1(u^*(t)s_x(t)) \right) \to 0. \quad (23)$$

Consider the following decomposition of the event of interest,

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

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

$$\mathbb{P} (Z_0(u^*(t)s_x(t)) + Z_1(u^*(t)s_x(t)) > f_{x,t}(u^*(t) + \epsilon/2)) \leq \frac{\mathbb{E}[Z_0(u^*(t)s_x(t)) + Z_1(u^*(t)s_x(t))]}{f_{x,t}(u^*(t) + \epsilon/2)} = \frac{f_{x,t}(u^*(t), t)}{f_{x,t}(u^*(t) + \epsilon/2)} = O(x^{-\lambda_1 \epsilon/2r}),$$

where the last equality follows from the ‘steepness’ at the minimum property (7).

We can do a further decomposition,

$$\mathbb{P} \left( \inf_{u \in [u^*(t) + \epsilon, \infty]} Z_1(us_x(t)) < f(u^*(t) + \epsilon/2) \right) = \mathbb{P} \left( \inf_{u \in [u^*(t) + \epsilon, \infty]} Z_1(us_x(t)) < f(u^*(t) + \epsilon/2), Z_1 ((u^*(t) + \epsilon)s_x(t)) < f_{x,t}(u^*(t) + 3\epsilon/4) \right) + \mathbb{P} \left( \inf_{u \in [u^*(t) + \epsilon, \infty]} Z_1(us_x(t)) < f(u^*(t) + \epsilon/2), Z_1 ((u^*(t) + \epsilon)s_x(t)) > f_{x,t}(u^*(t) + 3\epsilon/4) \right) \leq \mathbb{P} (Z_1 ((u^*(t) + \epsilon)s_x(t)) < f_{x,t}(u^*(t) + 3\epsilon/4)) + \mathbb{P} \left( \inf_{u \in [u^*(t) + \epsilon, \infty]} Z_1(us_x(t)) < f(u^*(t) + \epsilon/2), Z_1 ((u^*(t) + \epsilon)s_x(t)) > f_{x,t}(u^*(t) + 3\epsilon/4) \right) (24)$$

Using Chebyshev’s inequality and the result in (7) again, we see that

$$\mathbb{P} (Z_1 ((u^*(t) + \epsilon)s_x(t)) < f_{x,t}(u^*(t) + 3\epsilon/4)) = \mathbb{P} (Z_1 ((u^*(t) + \epsilon)s_x(t)) - \mathbb{E}[Z_1 ((u^*(t) + \epsilon)s_x(t))] < f_{x,t}(u^*(t) + 3\epsilon/4) - \mathbb{E}[Z_1 ((u^*(t) + \epsilon)s_x(t))]) \leq \frac{\mathbb{E} [Z_1 ((u^*(t) + \epsilon)s_x(t)) - \mathbb{E}Z_1 ((u^*(t) + \epsilon)s_x(t))]^2}{|f_{x,t}(u^*(t) + \epsilon/4) - \mathbb{E}Z_1 ((u^*(t) + \epsilon)s_x(t))|^2}. \quad (25)$$

29
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

$$
\frac{r}{\lambda_1} xe^{\lambda_0 u^*(t)\left(\frac{1}{\alpha} \log x + t\right)} \left( 1 - \frac{r_0 \mu}{x^\alpha (\lambda_1 + r)} \right) = \frac{x^{1-\alpha} r_0 \mu}{\lambda_1 + r} e^{\lambda_1 u^*(t)\left(\frac{1}{\alpha} \log x + t\right)}.
$$

Thus,

$$
\mathbb{E}[Z_1 ((u^*(t) + \varepsilon)s_x(t))] = xe^{\lambda_0 u^*(t)\left(\frac{1}{\alpha} \log x + t\right)} \left( 1 - \frac{r_0 \mu}{x^\alpha (\lambda_1 + r)} \right) e^{\lambda_1 \varepsilon\left(\frac{1}{\alpha} \log x + t\right)} + \frac{r}{\lambda_1} \left( 1 - \frac{r_0 \mu}{x^\alpha (\lambda_1 + r)} \right) e^{\lambda_1 \varepsilon\left(\frac{1}{\alpha} \log x + t\right)}.
$$

Also,

$$
f_x,t(u^*(t) + 3\varepsilon/4) = xe^{\lambda_0 u^*(t)\left(\frac{1}{\alpha} \log x + t\right)} \left( 1 - \frac{r_0 \mu}{x^\alpha (\lambda_1 + r)} \right) \left( e^{\lambda_0 3\varepsilon/4\left(\frac{1}{\alpha} \log x + t\right)} + \frac{r}{\lambda_1} e^{\lambda_1 3\varepsilon/4\left(\frac{1}{\alpha} \log x + t\right)} \right),
$$

and therefore

$$
|f_x,t(u^*(t) + 3\varepsilon/4) - \mathbb{E}Z_1 ((u^*(t) + \varepsilon)s_x(t))^2

\begin{align*}
&= \left( xe^{\lambda_0 u^*(t)\left(\frac{1}{\alpha} \log x + t\right)} \left( 1 - \frac{r_0 \mu}{x^\alpha (\lambda_1 + r)} \right) \left( e^{\lambda_0 3\varepsilon/4\left(\frac{1}{\alpha} \log x + t\right)} + \frac{r}{\lambda_1} e^{\lambda_1 3\varepsilon/4\left(\frac{1}{\alpha} \log x + t\right)} \right) \right)^2

&= \frac{r_0 \mu}{x^\alpha (\lambda_1 + r)} x^{\alpha}\left(\frac{\lambda_1}{\lambda_0} + 1 - u^*(t)\right) e^{\lambda_0\varepsilon\left(\frac{1}{\alpha} \log x + t\right)} + \frac{r}{\lambda_1} e^{\lambda_1\varepsilon\left(\frac{1}{\alpha} \log x + t\right)}

&= \Omega \left( x^{2\left(\frac{\lambda_1}{\lambda_0} + 1 - u^*(t)\right)} \right).
\end{align*}

Next we consider the variance term. For ease of notation define $\theta(t) \equiv (u^*(t) + \varepsilon)(\frac{1}{\alpha} \log x + t)$. Var[$Z_1 ((u^*(t) + \varepsilon)s_x(t))]$

$$
\begin{align*}
&= \left( \frac{r_0 \mu}{x^\alpha} \right)^2 \int_0^{\theta(t)} \int_0^{\theta(t)} \mathbb{E}[Z_0(s)Z_0(y)] x^\lambda_1 u^*(t)+\varepsilon/y e^{-\lambda_1 s x^\lambda_1 (u^*(t)+\varepsilon)/r} e^{-\lambda_1 y ds dy} \quad (i) \\
&+ \left( \frac{r_0 \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 \quad (ii) \\
&- \left[ \frac{r_0 \mu}{x^\alpha (\lambda_1 - \lambda_0)} x^{\lambda_1 (u^*(t)+\varepsilon)/r + 1} e^{\lambda_1 (u^*(t)+\varepsilon)t} \left( 1 - e^{(\lambda_0 - \lambda_1)(u^*(t)+\varepsilon)t} x^{(\lambda_0 - \lambda_1)(u^*(t)+\varepsilon)/r} \right) \right]^2 \quad (iii)
\end{align*}

Looking at the highest order term in (iii) we obtain:

$$
- \left[ \frac{r_0 \mu x^{1-\alpha}}{(\lambda_1 - \lambda_0)} x^{\lambda_1 (u^*(t)+\varepsilon)/r} e^{\lambda_1 (u^*(t)+\varepsilon)t} \right]^2 \quad (26)
$$

Clearly, the highest order term from (i) + (ii) is the highest order term from (i), which is:

$$
\frac{x^{2-2\alpha}(r_0 \mu)^2}{(\lambda_1 + r)^2} x^{2\lambda_1 (u^*(t)+\varepsilon)/r} e^{2\lambda_1 (u^*(t)+\varepsilon)t}
$$
which cancels with (26). From this we can conclude that

\[
\text{Var}[Z_1((u^*(t) + \varepsilon)s_x(t))] = o(x^{2-2\alpha}x^{2\lambda_1(u^*(t)+\varepsilon)/r}).
\]

Then, in order to establish that (25) goes to zero it suffices to show that

\[
\frac{o \left( x^{2-2\alpha+2\lambda_1} \frac{u^*(t)+\varepsilon}{r} \right)}{\Omega \left( x^{2(\frac{\lambda_1 \varepsilon}{2} + 1-u^*(t))} \right)} \to 0.
\]

Using the definition of \( u^*(t) \) in (6) we can establish that

\[
x^{2-2\alpha+2\lambda_1} \frac{u^*(t)+\varepsilon}{r}
\]

converges to a finite constant, and therefore we can conclude that

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

It remains to show that the final probability in display (24) 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 (24) 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

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

thus establishing (23).

Moving on we next establish that:

\[
\mathbb{P} \left( \inf_{u \in [0,u^*(t) - \varepsilon]} Z_0(us_x(t)) < Z_0(u^*(t)s_x(t)) + Z_1(u^*(t)s_x(t)) \right) \to 0 \quad (27)
\]
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 \( x \to \infty \):

\[
P \left( \inf_{u \in [0,u^*(t)-\varepsilon]} Z_0(us_x(t)) < f_{x,t}(u^*(t) - \varepsilon/2) \right)
\]

\[
= P \left( \inf_{u \in [0,u^*(t)-\varepsilon]} Z_0(us_x(t)) - cf_{x,t}(u^*(t) - \varepsilon/2) < (1 - c)f_{x,t}(u^*(t) - \varepsilon/2) \right)
\]

\[
\leq \frac{1}{(c-1)f_{x,t}(u^*(t) - \varepsilon/2)} E \left[ (cf_{x,t}(u^*(t) - \varepsilon/2) - Z_0((u^*(t) - \varepsilon)s_x(t)))^+ \right]
\]

\[
\leq \frac{cf_{x,t}(u^*(t) - \varepsilon/2)}{(c-1)f_{x,t}(u^*(t) - \varepsilon/2)} P \left( Z_0((u^*(t) - \varepsilon)s_x(t)) < cf_{x,t}(u^*(t) - \varepsilon/2) \right)
\]

where we chose \( c > 1 \), and the penultimate inequality follows from Doob’s inequality and that \(-Z_0(-s_x(t))\) is a submartingale. The final probability can be rewritten as (for \( x \) sufficiently large)

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

\[
E Z_0((u^*(t) - \varepsilon)s_x(t)) - cf_{x,t}(u^*(t) - \varepsilon/2) \geq C_0x^{1+\varepsilon-u^*(t)},
\]

and therefore

\[
P \left( Z_0((u^*(t) - \varepsilon)s_x(t)) \right) \leq x^{1/2} (\text{Var}_1 Z_0((u^*(t) - \varepsilon)s_x(t)))^{1/2}
\]

\[
\leq \frac{x^{1/2} (\text{Var}_1 Z_0((u^*(t) - \varepsilon)s_x(t)))^{1/2}}{C_0x^{1+\varepsilon-u^*(t)}}
\]

\[
= O \left( x^{-(1+\varepsilon-u^*)/2} \right),
\]

where the final equality follows by evaluating the variance of \( Z_0(u^*(t)(\frac{1}{t} \log x + t)) \). \( \blacksquare \)
7.6 Proof of Lemma 4

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

\[
\inf_{u \in [0,u^*(t_0)-\varepsilon]}\ Z_0(u(\frac{1}{r} \log x + t_0)) + Z_1(u(\frac{1}{r} \log x + t_0))
\]

\[
= \inf\{Z_0(s) + Z_1(s) : s \leq (u^*(t_0) - \varepsilon)(\frac{1}{r} \log x + t_0)\}
\]

\[
\geq \inf\{Z_0(s) + Z_1(s) : s \leq (u^*(t_1) - \varepsilon)(\frac{1}{r} \log x + t_1)\}
\]

\[
= \inf_{u \in [0,u^*(t_1)-\varepsilon]}\ Z_0(u(\frac{1}{r} \log x + t_1)) + Z_1(u(\frac{1}{r} \log x + t_1)).
\]

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 \( s, t \). Therefore if \( t_0 > t_1 \),

\[
\inf_{u \in [0,u^*(t_0)-\varepsilon]}\ Z_0(u(\frac{1}{r} \log x + t_0)) + Z_1(u(\frac{1}{r} \log x + t_0))
\]

\[
< Z_0(u^*(t_0)(\frac{1}{r} \log x + t_0)) + Z_1(u^*(t_0)(\frac{1}{r} \log x + t_0))
\]

implies

\[
\inf_{u \in [0,u^*(t_1)-\varepsilon]}\ Z_0(u(\frac{1}{r} \log x + t_1)) + Z_1(u(\frac{1}{r} \log x + t_1))
\]

\[
< Z_0(u^*(t_1)(\frac{1}{r} \log x + t_1)) + Z_1(u^*(t_1)(\frac{1}{r} \log x + t_1))
\]

so

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

as \( x \to \infty \). Then, the same argument can be used for deviations to the right of \( u^*(t) \) to show that

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

as \( x \to \infty \), establishing the lemma. ■

References


